TECHNICAL REPORT
NATICK/TR-09/004

AD ________________

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (DARPA) IMPROVING WARFIGHTER INFORMATION INTAKE UNDER STRESS: AUGMENTED COGNITION PHASES 2, 3, AND 4

by

Michael C. Dorneich
Patricia May Ververs
Santosh Mathan
and
Stephen D. Whitlow

Honeywell Laboratories
Minneapolis, MN 55418

November 2008

Final Report
June 2003 –January 2007

Approved for public release: distribution unlimited.

Prepared for
U.S. Army Natick Soldier Research, Development and Engineering Center
Natick, Massachusetts 01760-5056

DISCLAIMERS

The findings contained in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of trade names in this report does not constitute an official endorsement or approval of the use of such items.

DESTRUCTION NOTICE

For Classified Documents:

Follow the procedures in DoD 5200.22-M, Industrial Security Manual, Section II-19 or DoD 5200.1-R, Information Security Program Regulation, Chapter IX.

For Unclassified/Limited Distribution Documents:

Destroy by any method that prevents disclosure of contents or reconstruction of the document.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
18-11-2008	Final	June 2003 –January 2007

4. TITLE AND SUBTITLE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (DARPA) IMPROVING WARFIGHTER INFORMATION INTAKE UNDER STRESS: AUGMENTED COGNITION – PHASES 2, 3, AND 4

5a. CONTRACT NUMBER
DAAD-16-03-C-0054

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael C. Dorneich, Patricia May Ververs, Santosh Mathan, and Stephen D. Whitlow

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Honeywell Laboratories
3660 Technology Drive
Minneapolis, MN 55418

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Natick Soldier Research, Development and Engineering Center
Kansas St., ATTN: AMSRD-NSR-TS-A (H. Girolamo)
Natick, MA 01760-5056

10. SPONSOR/MONITOR'S ACRONYM(S)
NSRDEC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NATICK/TR-09/004

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report is a comprehensive summary of a multi-year effort by the Honeywell team on the Improving Warfighter Information Intake Under Stress/AugCog program jointly sponsored by the Defense Advanced Research Project Agency (DARPA) and the U.S. Army. The team, which spanned industry, government, and academia, studied the measurable cognitive states of the dismounted Soldier. The first seven months of Honeywell's involvement consisted of studies that developed neurophysiological and physiological measures of cognitive states, particularly attention. The next two years of the program focused on the challenges of assessing the cognitive state of a mobile participant and the development of mitigation strategies to improve the overall throughput of the joint human-machine system. The final year's effort proved the feasibility of the AugCog technology for the dismounted Soldier by testing the system in a military Mobile Operations in Urban Terrain (MOUT) environment with a platoon of Soldiers. The Honeywell team believes it was the first ever to demonstrate robust real-time cognitive state classification in the harsh operational MOUT environment. The classification accuracies obtained in the final study match those of the more pristine laboratory environment despite the motion, noise, and physical challenges posed by collecting physiological data in the field during real operations.

15. SUBJECT TERMS
AUG-COG(AUGMENTED COGNITION) STRESS(PHYSIOLOGY) PHYSIOLOGICAL MEASUREMENT
REAL TIME ECG COGNITIVE STATES STRESS(PSYCHOLOGY)
ARMY PERSONNEL EEG DECISION MAKING SENSORS
INFORMATION FLOW PERFORMANCE(HUMAN) WORKLOAD
MOUT(MILITARY OPERATIONS ON URBAN TERRAIN) MITIGATION STRATEGIES

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Henry Girolamo
U	U	U	SAR	232	19b. TELEPHONE NUMBER *(include area code)* (508) 233-5483

INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

Preface

The Defense Advanced Research Project Agency (DARPA) Improving Warfighter Information Intake Under Stress (IWIIUS)/Augmented Cognition (AugCog) program was a four-year, four-phase program. Honeywell participated in the last three of the four phases, from June 2003 to January 2007, under contract (number DADD16-03-C-0054) to Natick Soldier Research, Development and Engineering Center (NSRDEC). Work in the field of Augmented Cognition began by establishing the ability to classify, in real-time, cognitive processing states (attention, working memory, executive function, and sensory memory) with laboratory tasks (known as Psych 101 tasks). Phase 1 of the program concentrated on developing technologies that could measure cognitive states, via brain imaging, external brain monitoring, body sensing, and eye measures. Gradually over the past three years, researchers have moved from the laboratory environment to the field environment, introducing the artifacts (e.g. motion, electrical, networking traffic and disconnects) and stressors (e.g., information overload, physical load, competition, threat of pain) inherent in the operational environment to which the technology would be transitioned. This report details the research conducted in cognitive state assessment, the development of closed-loop integrated prototypes, and evaluation findings obtained by the Honeywell team in each of the last three phases of the program.

From the start of the program, the Honeywell AugCog team worked closely with the U.S. Army to address the problem of information overload that is expected to occur with the rapid deployment of Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) technologies. In the next decade, unparalleled information sharing and real-time collaboration across geographically diverse assets will occur and impact the individual Soldier. When deployed correctly, these technologies will provide greater situational understanding for decisive actions. However, success will be dependent on the Warfighter's ability to sort through the vast array of continuous information flow afforded by a full range of netted communications. The Army recognizes the potential strain that added capabilities will impose on deployed Soldiers operating in the stressful conditions of war. Therefore, as new systems are spun into the Army's Ground Soldier System program, requirements exist for systems to be developed to assist Soldiers during all operational conditions, particularly when the Soldiers' cognitive skills are degraded. The first step is recognizing when these degraded cognitive states exist. Augmented Cognition technologies developed during the four-year program offer that ability to detect degraded cognitive states.

Throughout the program, the Honeywell team benefited from the contributions of multiple subcontractors. The Honeywell team consisted of Honeywell Laboratories and, at various phases, Advanced Brain Monitoring, Carnegie Mellon University, City College of New York, Clemson University, Columbia University, Drexel University, Human Bionics, Institute of Human and Machine Cognition (IHMC), Oregon Health and Sciences University, Sarnoff, UFI, University of New Mexico, and the University of Virginia.

Acknowledgments

Program Management

The Honeywell team would like to thank the DARPA management team, Dr. Amy Kruse, CDR Dylan Schmorrow, Ph.D., and (Ret.) Admiral Lee Kollmorgen for their programmatic guidance in directing the AugCog program. In addition, we would like to thank Colby Raley and Ami Bolton for their program support. We would also like to acknowledge the steadfast support and guidance from Mr. Henry Girolamo as the DARPA Agent, and Dr. Jim Sampson for ensuring our efforts had operational relevance to the Future Force Warrior.

Management Support

The Honeywell team would like to acknowledge the support and guidance of Dr. Bill Rogers, Ms. Barbara Brockett, and Ms. Rose Mae Richardson within Honeywell Labs.

Operational Experiment

Over the four years of Honeywell involvement within the Augmented Cognition program, we have had the pleasure of collaborating with a wide range of business, government, and academic partners.

First, we would like to acknowledge the support of the Natick Soldier Research, Development and Engineering Center (NSRDEC) and the continual support of Mr. Henry Girolamo and Dr. James Sampson. Mr. Dennis Magnifico of the Army Technology Transition Office was invaluable in coordinating activities with the various Army offices. We would like to thank Dr. Caroline Mahoney for her assistance with the Hidalgo Vital Signs Detection System lab testing. The technology transition demonstration to the Army would not have been possible without the support and guidance of Ms. Cynthia Blackwell and Mr. Adam Malhoit of the Future Force Warrior Program.

We would like to thank our collaborators at Carnegie Mellon University (Dr. Randy Pausch), Clemson University (Dr. Eric Muth, Dr. Adam Hoover), Columbia University (Dr. Paul Sajda), City College of New York (Dr. Lucas Parra), Human Bionics (Don DuRousseau), Institute of Human and Machine Cognition (Dr. Anil Raj), Office of Naval Research (LT Joseph Cohn, Dr. Roy Stripling), Oregon Health and Sciences University (Dr. Misha Pavel, Dr. Tamara Hayes, Denis Erdogmus, A. Adami, L. Tan), UFI (Marty Loughry), University of New Mexico (Dr. Akasha Tang), and University of Virginia (Dr. Denny Proffitt).

During the early phases, we relied on two organizations to host the concept validation experiments (CVEs): IHMC and Carnegie Mellon University. We would like to thank Mr. Roger Carff and Mr. Matt Johnson for their work on the agent architecture to enable

the sensor integration and Mr. Jeremy Higgins for his assistance in running the participants at IHMC. We would also like to acknowledge the efforts of Dr. Randy Pausch, the CMU graduate students—Allison Styer, Rob Gordon, Mike Darga, and Kyle Gabler directed by Jesse Schell, who were responsible for the many iterations in the Panda3D virtual environment and associated tasks, and Jason Pratt and Ben Buchwald, who were primarily responsible for interfacing the Panda environment with the IHMC architecture and the data collection. We would also like to thank Jessica Hodgins for the use of the Motion Capture laboratory during the Concept Validation Experiment (CVE) and the periodic demonstrations.

The Phase 4 Operational Experiment was a tremendous undertaking that involved seven organizations. The Honeywell team would like to acknowledge the efforts of the multiple organizations that came together to make the Augmented Cognition Test Event a success.

As throughout the AugCog program, we received continual support and guidance from the NSRDEC. We would like to thank Mr. Henry Girolamo, Dr. Jim Sampson, and Mr. Dennis Magnifico for their guidance in developing the operational scenarios, assistance in the field, and support throughout the test event. We would also like to thank the Battle Lab Integration Team (BLIT), in particular, Mr. Fred Dupont, for taking the lead on developing an Army exercise that would both meet the Soldier's training needs while allowing the Honeywell team to meet their experiment objectives. In addition, Fred led the two weeks of training during the Augmented Cognition Test Event (ACTE). Fred's tireless devotion to ensuring that everyone got what they needed was critical in making the ACTE a success for all involved. We would also like to thank Dr. Ken Parham and Mr. Chris King for their help in Soldier coordination and Soldier training during the ACTE.

We would like to thank the Aberdeen Test Center (ATC), particularly Mr. Tony Ham, who acted as the overall test coordinator, led the submission of the test plan, and coordinated with onsite Soldiers for support. In addition, we would like to thank Mr. Paul Tennant, Ms. Reta Reynolds, and Mr. Jim Buxton for assistance with the Soldiers.

We would also like to thank USARIEM, especially Mr. Bill Tharion, for experimental investigation (human use) as well as serving as a shadower for six days; Mr. Mark Buller for acting as Hidalgo system liaison; Mr. Steve Mullen and Mr. Tony Karis for system engineering support; and Col. Beau Freund and Maj. Latzka of the Warfighter Physiological Status Monitoring program.

We would like to thank Hidalgo Inc., in particular, Mr. Justin Pisani (CEO), for coordination on integrating the Vital Signs Detection System.

We would like to thank the Human Research and Engineering Directorate (HRED), in particular, Dr. Scott Kerick, for consultations on EEG in the field.

We would like to thank the Development Test Command (DTC), particularly Mr. Bob Gauss, for assistance with the safety release, Mr. Jorge Hernandez for assistance with the Human Use Research Committee (HURC) protocol, and Dr. Dal Nett, the HURC chair.

The Honeywell team would like to express its profound gratitude to the Soldiers of the North Carolina Army National Guard, 1/252nd Combined Arms Battalion, stationed at Ft. Bragg, N.C. Their professionalism, dedication to training, and adaptability in handling the extra demands imposed by the experiment were vital in making the ACTE a success.

Finally, we would like to acknowledge the efforts of the Honeywell research and development team that supported the development of the prototypes and the running of all experiments. We would like to make special mention of the efforts of Ms. Danni Bayn, Dr. Kelly Burke, Mr. Jim Carciofini, Ms. Janet Creaser, Mr. Bob DeMers, Mr. Trent Reusser, and Mr. Jeff Rye.

Executive Summary

The Defense Advanced Research Projects Agency (DARPA) Improving Warfighter Information Intake Under Stress (IWIIUS)/Augmented Cognition (AugCog) program was a four-year, four-phase program. Honeywell participated in the last three of the four phases from June 2000 to January 2007. Phase 1 of the program concentrated on developing technologies that could measure cognitive state, via brain imaging (e.g., functional Near Infrared (fNIR)), external brain monitoring (e.g., Electroencephalogram (EEG)), body sensing (e.g., Electrocardiogram (ECG) based arousal), and eye measures (e.g., pupillary reflexes).

From the start of the Phase 2 program, the Honeywell AugCog team worked closely with the U.S. Army to address the problem of information overload expected to occur with the rapid deployment of C4ISR (Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance) technologies. In the next decade, unparalleled information sharing and real-time collaboration across geographically diverse assets will occur and impact the individual Soldier. When deployed correctly, the technologies will provide greater situational understanding for decisive actions. However, success will be dependent on the Warfighter's ability to sort through the vast array of continuous information flow afforded by a full range of netted communications. The Army recognizes the potential strain the added capabilities will impose on deployed Soldiers operating in the stressful conditions of war. Therefore, as new systems are spun into the Army's Ground Soldier System program, requirements exist for systems to be developed to assist Soldiers during all operational conditions, particularly when the Soldiers' cognitive skills are degraded. The first step is recognizing when these degraded cognitive states exist. Augmented Cognition technologies offer that ability to detect degraded performance states.

Phase 2 of the AugCog program was an 18-month effort that began in June of 2003. The Honeywell team consisted of Honeywell Laboratories, Carnegie Mellon University, Clemson University, Columbia University, Sarnoff, Human Bionics, Institute of Human and Machine Cognition (IHMC), Oregon Health and Sciences University, UFI, University of New Mexico, and University of Virginia. The phase was segmented into two parts: Phase 2a, a six-month effort, and Phase 2b, a 12-month effort. Phase 2a focused on the manipulation and measurement of cognitive state in general and, for Honeywell, attention in particular. The DARPA phase metrics included the ability to detect cognitive state shifts in less than two seconds and trigger a cognitive state manipulation (mitigation) within one minute of the onset of a cognitive state shift. Specifically, in Phase 2a, Honeywell concentrated on developing a closed-loop system (CLIP) that triggered information management mitigations to reduce demands on attention and improve participant performance. Cognitive workload assessment was driven by a comprehensive suite of sensors, including electroencephalogram (EEG), pupilometry, electrodermal response (EDR), electrocardiogram (ECG), and electromyogram (EMG). These sensors served as inputs to five cognitive state gauges (Arousal Meter, Stress Gauge, Engagement Index, eXecutive Load Index (XLI), and P300-driven novelty detector), as well as more straightforward measures of physiology

such as heart rate and Interbeat Interval (IBI) from the ECG. The participants' interaction with the CLIP was evaluated in a virtual environment (VE) with tasks that approximated the cognitive load of military operational tasks (identification of friend or foe, engagement of foes, navigation, and communications). There were significant trial-wide and within-trial effects from the IBI measure indicating higher IBIs on Augmented trials compared to non-Augmented trials. This suggests that the Augmentation intervention decreased participants' autonomic arousal. The findings also indicated that the XLI accurately differentiated between low-load and high-load conditions without Augmentation in 10 out of 11 participants. Overall, the findings indicated significant positive correlations between average gauge correlations for all participants, indicating that not only were there redundant measures that were sensitive to experimental manipulation, but that they detected both neurophysiological and physiological responses to task load.

The Phase 2b effort involved Honeywell Laboratories, Carnegie Mellon University, City College of New York, Clemson University, Columbia University, Human Bionics, Institute of Human and Machine Cognition, Oregon Health and Science University, and UFI. In Phase 2b, Honeywell conducted two separate Concept Validation Experiments (CVEs). The first CVE was held at IHMC and focused on the development of mitigation strategies with military-relevant tasks (such as navigate to an objective, engage foes, and attend to radio communications) performed in the virtual environment used in Phase 2a. The second CVE, held at Carnegie Mellon University's Motion Capture (MoCap) laboratory, focused on the ability to detect cognitive state in a (semi-) mobile virtual environment. These environments were chosen because of the flexibility they offered in creating scenarios that were operationally realistic. These environments also provided the ability to manipulate the attentional demands associated with tasks. Situating tasks within these virtual environments allowed experimenters to precisely relate simulation events to neurophysiological states assessed by the gauges. The two virtual environments also provided insight into the performance of the gauges under different levels of mobility.

The findings of the Phase 2b study conducted at IHMC indicated several significant performance improvements, including a 100% improvement in message comprehension and a 125% improvement in situation awareness with the Communications Scheduler mitigation. There was a 380% decrease in the number of ambushes encountered with the tactile navigation cueing mitigation. There was a 96% improvement in the communication of critical information with the Medevac Negotiation Assistance. The CMU CVE focused on improving overall performance involved with the identification of friend or foe and radio communications tasks in a mobile environment. A scheduling mitigation was applied to the secondary task of radio communications in which the participants needed to maintain a running count of reported friendlies and enemies spotted by team leaders in working memory. If the gauges detected a high workload condition, the mitigation deferred radio messages until after the completion of the primary task. The gauge-based scheduling strategy produced a 60% improvement in performance and significantly lowered perceived mental workload in the mitigated condition as compared to the unmitigated (random scheduling) condition.

Phase 3 was a 12-month effort, and the team consisted of Honeywell Laboratories, Advanced Brain Monitoring, Inc (ABM), Oregon Health and Science University, and

Drexel University. An evaluation was conducted to investigate the efficacy of two mitigation strategies (a communications scheduler and a tactile cueing mitigation) outside the laboratory in a wooded field environment. The communications scheduler mitigation was driven by an assessment of the participant's current cognitive capacity to process incoming information, and scheduled the presentation of communications in order to improve decision making under high task load conditions. A tactile cueing mitigation was created to support the participant's navigation along a complex route when competing tasks drove cognitive workload too high. The evaluation was conducted to demonstrate whether cognitive capacity to perform under differing task loads could be detected using neurophysiological sensors and if adaptive automation mitigations would appropriately regulate information flow. The communications scheduler resulted in an improvement in primary task performance (a maintain counts working memory task) and lower subjective workload ratings without a degradation in concurrent secondary tasks (mission monitoring and math tasks). The tactile cueing mitigation provided non-visual navigation support that offloaded a typically visual task, such as reading a paper map or computer-based map display. Under the mitigated condition using the tactile cueing, a decreased runtime on the 'navigate to objective' task during the high task load period was found. The findings also revealed the need for caution when applying automated support for navigation due to potential costs to the situation awareness of surroundings. The results suggest that the automation should only be used in high-workload situations where the benefits outweigh these costs.

Also in Phase 3, the Honeywell team worked with the Army at Aberdeen Proving Ground, Aberdeen Test Center (ATC) to demonstrate the cognitive state assessment of a Commander performing in an operational exercise. At the Joint Distributed Freeplay Event (JDFE) at Mulberry Point Military Operations in Urban Terrain (MOUT) site, the Honeywell technology was used to assess the cognitive state of the Joint Task Force (JTF) Commander. EEG data collected in the context of the JDFE event provided an opportunity to assess the potential for Honeywell's real-time classification approach in an operationally relevant task environment. The premise of the event centered on a joint personnel recovery mission in which a downed pilot was captured by enemy insurgents and a rescue mission was planned and executed. The AugCog team outfitted the JTF Commander with a six-channel wireless EEG cap manufactured by ABM integrated into Honeywell's information architecture. The primary role of the JTF Commander was to maintain communications with the JTF staff to gather intelligence regarding movements of the opposing force and support the blue force (BLUFOR) squad leader leading the recovery mission in the field. Using the variations in the cognitive workload required of the scenario, the Honeywell AugCog team evaluated the classification techniques previously used in the laboratory and field tests to classify cognitive state. EEG data was collected from three different commanders and submitted to the refined classification approach. Classification results varied from a low of 65% accuracy to a high of 78% accuracy. The Phase 3 section discusses how these promising results laid the groundwork for future refinements in the classification methodology.

The final 12-month phase, Phase 4, focused on operational feasibility. The team consisting of Honeywell Laboratories, Oregon Health and Science University, and ABM worked closely with the U.S. Army to evaluate a streamlined and refined AugCog system

with a platoon of Soldiers. The culminating test event to the IWIIUS/AugCog program was an evaluation in a MOUT environment at the Aberdeen Proving Ground performed over a two-week period. The overall objective was to assess Soldier workload levels during various operational tasks requiring different levels of cognitive and physical engagement, and demonstrate the effectiveness of the AugCog techniques as measures of cognitive loading during mission phases on key leadership positions. The team evaluated the effectiveness of sensor-driven assessment of cognitive state, looking at both physiological (ECG) and neurophysiologically (EEG) based sensors. For the first time in the program the Honeywell AugCog team explored a sensor fusion approach to cognitive state classification. Fusing the brain measures from EEG with cardiac data enabled a substantive boost to overall classification performance, resulting in classifications of 86% and 95% of two leaders using a tenfold cross-validation training approach. These workload classification accuracies match those obtained in more controlled laboratory environments despite the motion, noise, and physical challenges imposed by collecting physiological data in the field during real operations.

The Honeywell AugCog team is well positioned to leverage the advancements made on the DARPA/Army-sponsored program and transition the advanced technologies to the next-generation Soldier systems being developed by the Army. A follow-on effort with the Army will be focusing on evaluating the Honeywell AugCog system's capability to enable remote Army Commanders and other Army leaders to assess the cognitive state assessment of dismounted Soldiers in real-time during a battlefield exercise. This system will further enable improvements in military operational decision making required by the Army by providing dynamically updated Soldier readiness gauges. The validation of the refined system will take place in a Future Force Warrior (FFW) operational demonstration in 2007. The Honeywell program will deliver and evaluate the efficacy of this real-time, wireless, wearable solution that will use single-trial EEG spatial frequency patterns and ECG measures of IBI and heart rate variability to construct the real-time cognitive state assessments.

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (DARPA) IMPROVING WARFIGHTER INFORMATION INTAKE UNDER STRESS AUGMENTED COGNITION PHASES 2, 3, AND 4

1 Introduction

This report is a comprehensive summary of a nearly four-year effort (from June 2003 to January 2007) by the Honeywell team on the Improving Warfighter Information Intake Under Stress(IWIIUS)/Augmented Cognition (AugCog) program, Honeywell's efforts were jointly sponsored by Defense Advanced Research Projects Agency (DARPA) and the U.S. Army, under contract to Natick Soldier Research, Development and Engineering Center (NSRDEC). In a team effort that spanned industry, government, and academia, Honeywell set out to study the measurable cognitive states of the dismounted Soldier. The first six months of Honeywell's participation consisted of studies that developed neurophysiological and physiological measures of cognitive states, particularly attention. The next two years of the program focused on the challenges of assessing the cognitive state of a mobile participant and the development of mitigation strategies to improve the overall throughput of the human-machine system. The final year proved the feasibility of the AugCog technology for the dismounted Soldier by testing the system in a military Mobile Operations Urban Terrain (MOUT) environment with a platoon of Soldiers.

1.1 Operational Environment of the Dismounted Soldier

The U.S. Department of Defense (DoD) has embarked on a process of change called Transformation to create a highly responsive, networked, joint force capable of making swift decisions at all levels and maintaining overwhelming superiority in any battle space (Parmentola, 2004). In response, the U.S. Army is shaping its Future Force to be smaller, lighter, faster, and smarter than its predecessor. The network will be characterized by a network of humans collaborating through a system of Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) technologies.

Evidence of the Army Transformation could already be seen in Operations Enduring Freedom and Iraqi Freedom. Some of the most visible and valuable benefits were seen in the speed of operations, enabling reduction in the time to plan missions, make decisions, and coordinate and move large groups of Soldiers. What was created was a more dynamic and adaptive operation built on the collective capabilities of all the participants. Unprecedented levels of integration took place among the air, naval, and land forces. Stone (2003) reports, for instance, that in the middle of Afghanistan, special operations Soldiers could link with a Navy F-14 or link with a B-52 to pursue a target. As a perfect example of the creative innovation of the Transformation, Wolfowitz, (2002) writes that "[s]pecial Forces on the ground have taken 19th century horse cavalry, combined it with

50-year-old B-52 bombers, and, using modern satellite communications, have produced truly 21st century capability”.

One of the core capabilities of the Transformation is the availability of netted communications enabling information sharing and real-time collaboration enhancing the kind of situational understanding that drives decisive actions. The Future Force Warrior will have unparalleled connectivity to build situation awareness right down to the individual Soldier. Mission success will be dependent on the individual Warfighter’s ability to sort through the vast array of continuous information flow afforded by the full range of netted communications.

The research described in this comprehensive report was aimed at validating the applicability of non-invasive neurophysiological and physiological state detection techniques for dismounted Soldier combat operations. The Honeywell Augmented Cognition (AugCog) team set out to develop warfighting concepts that could substantially increase the combat effectiveness of infantry small combat units. The objective was to enhance human performance and improve survivability through more effective Soldier readiness assessment and more effective information management. This could only be done with improved overall situation awareness from the top of the command down to the adaptable small units and individual Soldiers.

1.2 Foundations of Augmented Cognition

Considering the projected information processing load of future Warfighters, it seemed reasonable to propose automation solutions to help better manage their workload. However, automated solutions come with inherent risks, and the pros and cons of automating complex systems have been widely discussed in the literature (e.g., Parasuraman & Miller, 2004; Sarter, Woods & Billings, 1997). Automated systems bring precision and consistency to tasks, relieve operator monotony and fatigue, and contribute to economic efficiency. However, as widely noted, poorly designed automation can impose several undesirable consequences. Automation can relegate the operator to the status of a passive observer—serving to limit situation awareness—and induce cognitive overload when a user may be forced to inherit control from an automated system.

Some researchers have proposed that some of these negative consequences can be eliminated by designing automated systems that have traditionally adapted based on task, context, and performance models (Hancock & Chignell, 1987; Parasuraman et al., 1992; Scerbo, 1996). Adaptive automation, where the automation adapts during execution to the current task environment, can either provide adaptive aiding, which makes a certain component of a task simpler, or can provide adaptive task allocation, which shifts an entire task from a larger multitask context to automation (Parasuraman, Mouloua, & Hilburn, 1999). Adaptive systems must make timely decisions on how best to use varying levels of (adaptive) automation to provide support in a joint human-automation system. In order for an adaptive system to decide when to intervene, it must have some model of the context of operations, be it a functional model of system performance, or possibly a model of the operator’s functional state. Currently, many adaptive systems derive their inferences about the cognitive state of the operator from mental models, performance on the task, or external factors related directly to the task environment (Wickens & Hollands, 2000). For example, Scott (1999) developed the Ground Collision-Avoidance

System (GCAS) for test on an F-16D. GCAS used the projected time until an aircraft broke through a pilot-determined minimum altitude as an external condition to infer that a pilot's attention was incapacitated, at which point the system would perform a "fly up" evasive maneuver to avoid a ground collision. In that case, the automation took over control of the aircraft from the pilot.

There are a number of limitations of model-based adaptive systems. First, in many task environments it is impractical to instrument the system in order to infer cognitive load from overt behavior. Second, task demands change in unpredictable ways in many complex task environments in the military. Third, users respond to task demands as a function of prior experience. Fourth, model creation in complex task domains is very time consuming and expensive, and these recurring costs grow with the complexity of environment and variability of individuals.

In response to this identified opportunity, DARPA launched the AugCog program, later named the Improving Warfighter Information Intake Under Stress program. The aim of AugCog was to use physiological and neurophysiological sensors to detect states where human cognitive resources may be inadequate to cope with mission relevant demands. The goal was to enhance human performance when task-related demands surpassed the human's current cognitive capacity, which fluctuated subject to fatigue, stress, overload, or boredom. Efforts focused on ways to leverage cognitive state information to drive adaptive systems to manage information flow when detected human cognitive resources were inadequate for the tasks at hand (Dorneich, Ververs, Mathan, Whitlow, et al., 2006).

Neurophysiologically and physiologically triggered adaptive automation offers many advantages over the more traditional approaches to automation by basing estimates of operator state on directly sensed data. These systems offer the promise of leveraging the strengths of humans and machines by augmenting human performance with automation specifically when assessed human cognitive capacity falls short of the demands imposed by task environments. With more refined estimates of the operator's cognitive state, measured in real-time, adaptive automation also offers the opportunity to provide aid even before the operator knows he or she is getting into trouble. This approach does not require instrumentation of systems to record behavioral actions required for task model-based systems.

Seminal research in this area considered electroencephalographs (EEG) to assess operator mental workload. Specifically, work by Pope, Bogart, & Bartolome (1995) used the Engagement Index, a ratio of EEG power bands (beta/(alpha + theta)) as a measurement of how cognitively engaged a person was in a task, to trigger automation adaptation designed to maintain an optimum workload level throughout task execution. Subsequent studies replicated the findings and demonstrated how an adaptive system, based on the Engagement Index, could be used to improve performance compared to an unaided condition (Freeman et al., 1999; Mikulka, Scerbo, & Freeman, 2002).

Honeywell's participation in the DARPA program began in phase 2, in June 2003. During that phase the Honeywell AugCog team implemented a version of the Engagement Index to be used as a trigger for Honeywell's adaptive automation prototype that managed incoming communications traffic based on sensed cognitive state. However, results from the Engagement Index suggested that it could not be relied on to

provide the moment-to-moment classification accuracy required by the communication scheduler (Whitlow, Dorneich, Ververs, Raj, et al., 2004). There were a number of reasons why the Engagement Index was not a reliable trigger within the testing environment. First, participants were engaged in an immersive virtual combat game that was much more heterogeneous and less controlled than those tasks used in previous laboratory studies. Accordingly, the Honeywell team had less control in manipulating the participants' cognitive state and thus expected greater idiosyncratic responses patterns at the neural level. Second, unlike previous studies, the team could not rely on across-trial averaging to find differential responses from the Engagement Index. The virtual task environment and the communication scheduling prototype required near-real-time assessment of cognitive state and response in order to improve communications management. Therefore, this application required an adaptive automation trigger that had very high moment-to-moment classification accuracy that would respond to more general cognitive states to avoid the more limited applicability of measures such as the Engagement Index. There were a number of factors that required this investigative technique:

1) *Individual Differences*. As Scerbo et al. (2001) pointed out, there are unique individual EEG responses to task demands. While the characterization of the relationship between engagement and EEG activity in terms of activity within certain frequency bands and sites is useful for synthesizing broadly observed trends, a given individual's responses may deviate substantially from assumptions derived from averaged data. In response, some researchers have called for an approach that is more sensitive to individual variability in EEG expression (Mathan et al., 2005).

2) *Linear Relationships*. The Engagement Index was based on a linear relationship between power estimates at specific frequency bands. However, there are potentially informative nonlinear relationships across spectral features at various sites that could help discriminate between various cognitive states. Research indicated that more advanced pattern recognition techniques, such as multilayer neural networks, could exploit relationships among features that do not conform to linearity assumptions (Scerbo et al., 2001; Wilson & Russell, 2003).

3) *Analysis Windows*. The Engagement Index was designed to estimate cognitive state over an analysis window that was close to a minute in duration. Developers of the Engagement Index made no claims about its efficacy at temporal resolutions of a few seconds, or hundreds of milliseconds. In the authors' own laboratory experience, the Engagement Index was reliably able to discriminate between periods of high-intensity virtual combat and periods of rest in a first-person video game over the course of analysis windows that spanned minutes, but not at a resolution of less than 10 seconds (Whitlow et al., 2004). The demands of the task environment may require techniques that provide reliable cognitive state estimates with a fairly high degree of temporal resolution.

4) *Validation Context*. Much of the literature associated with cognitive state estimation relies on findings from data collected in relatively stationary laboratory settings (Schmorrow & Kruse, 2002). Data collection in laboratory environments has several attributes that cannot be realized in mobile contexts. For example (a) the experiment setup can be controlled in order to facilitate better performance, (b) various precautions to improve signal quality can be implemented, and (c) large-scale data collection, analysis,

and signal processing hardware and software can be used. These constraints have to be relaxed in mobile environments. In mobile applications, EEG signals can be very noisy and can be contaminated by a wide range of noise artifacts. Furthermore, the system must be portable and able to work in real time.

Many of these concerns are not unique to the Engagement Index. Other indices such as Arousal Meter (Hoover & Muth, 2004), ABM Workload and Vigilance gauges (Berka et al., 1999), and fNIR based cognitive state assessment (Izzetoglu et al., 2004) suffer from similar limitations.

In Phase 3, the Honeywell team addressed some of the shortcomings highlighted above by creating a system that was optimized to the unique EEG spectral characteristics of each individual in response to specific task demands. Pattern recognition techniques that make no restrictive assumptions about the form of the data being modeled were used. The system provided cognitive state estimates at a high degree of temporal resolution and was designed to work in real time in mobile contexts.

Three aspects of the approach are highlighted in the pages that follow: hardware integration into a wireless wearable form factor, real-time signal processing to detect and correct for noise artifacts, and a nonlinear classification approach.

Realizing the vision of an augmented cognition system in the context of an ambulatory Soldier has been constrained by several challenges. First, as Schmorrow and Kruse (2002) noted, processing and analysis of neurophysiological data have been largely conducted off-line by researchers and practitioners. However, in order for augmented cognition technologies to work in practical settings, effective and computationally efficient artifact reduction and signal processing solutions are necessary. Second, inferring the cognitive state of users demands pattern recognition solutions that are robust to noise and the inherent nonstationarity in neurophysiological signals (Popivanov & Mineva, 1999). Third, understanding the fluctuations of cognitive state in applied environments requires the development of means to collect reliable neurophysiological data outside the laboratory. Fourth, experiments must be designed, often under conflicting constraints (e.g., operationally realistic tasks vs. well-understood, controlled laboratory tasks), to effectively evaluate classification accuracy. Finally, compact and robust form factors (e.g., size, weight, ruggedness) associated with neurophysiological sensors and processors are a matter of critical concern.

1.2.1 Real-Time Signal Processing Challenges

Conducting military maneuvers in operational environments such as urban terrain often does not allow an individual to remain stationary and can demand simultaneous cognitive and physical activity. Consequently, difficulties related to processing of EEG signals in real-world settings include factors associated with both participant motion and the operational environment itself. Thus, utilization of research methods involving EEG in operational environments necessitates the use of real-time algorithms for signal detection and removal of artifacts. Although real-time signal processing and classification of the EEG has been implemented previously (Gevins & Smith, 2003; Berka, Levendowski, Cvetinovic, Petrovic, et al., 2004), it has not been realized in a truly mobile, ambulatory environment.

Inferring cognitive state from noninvasive neurophysiological sensors is a challenging task even in pristine laboratory environments. High-amplitude artifacts ranging from eye blinks to muscle artifacts and electrical line noise can easily mask the lower amplitude electrical signals associated with cognitive functions. These concerns were particularly pronounced in the context of ongoing efforts to realize neurophysiologically driven adaptive automation for the dismounted ambulatory Soldier. In addition to the typical sources of signal contamination, mobile applications must consider the effects of artifacts induced by shock, cable movement, and gross muscle movement. Specifically, artifacts related to participant motion include high-frequency muscle activity, verbal communication, and ocular artifacts consisting of eye movements and blinks; whereas artifacts related to the operational environment include instrumental artifacts such as electrical noise that created interference with the EEG signal (c.f. Kramer, 1991).

1.2.2 Classification Challenges

The use of EEG as the basis for cognitive state assessment was motivated by characteristics such as good temporal resolution, low invasiveness, low cost, and portability. While EEG offered several benefits, there were shortcomings related to the noise artifacts described above and the nonstationarity of the neural signal pattern over time. Despite these challenges, research has shown that EEG activity can be used to assess a variety of cognitive states that affect complex task performance. These included working memory (Gevins & Smith, 2000), alertness (Makeig & Jung, 1995), executive control (Garavan, Ross, Li, & Stein, 2000), and visual information processing (Thorpe, Fize, & Marlot, 1996). These findings pointed to the potential for using EEG measurements as the basis for driving adaptive systems that demonstrate a high degree of sensitivity and adaptability to human operators in complex task environments.

1.2.3 Scenario Design Challenges

In addition to the practical and system configuration challenges faced when moving from the laboratory to field studies, there are issues of experiment control and the characterization of cognitive state in less constrained environments. It is essential to select tasks that are both operationally relevant and afford reasonable adaptations that improved performance. In the laboratory, it is possible to develop simple tasks where workload is manipulated precisely and consistently. Additionally, a user's performance can be collected and evaluated accurately. This makes it relatively easy to establish ground truth about a user's likely workload. However, when developing operationally relevant tasks in a field environment, it becomes substantially harder to manipulate workload precisely and to interpret and assess a user's performance without compromising operational realism. The mobile field evaluation reported herein had two objectives. The first objective was to determine whether an operationally relevant task load manipulation had a measurable impact on a user's workload. The second objective was to establish whether a sensor-based classification approach could effectively classify a user's workload in a mobile setting.

1.2.4 Limitations: Long-Term Generalization

While results presented in this report suggest that robust and accurate classification is feasible in the field, a qualitative analysis of longitudinal data spanning days suggests that

much more research is necessary to create classifiers that can generalize over time spans of days as the task context and patterns of general physiological activity change.

1.3 Program Research Approach

The research done to address many of the challenges described above focused on two parallel and complementary thrusts: cognitive state assessment and mitigations. Phase 1 of the program was dedicated to proving that it was possible to reliably determine cognitive state in real-time based on brain imaging (e.g., fNIR), external brain monitoring (e.g., EEG), body sensing (e.g., electrocardiogram (ECG)-based arousal), and eye measures (e.g., pupillary reflexes). Phases 2-4 then closed the loop by driving adaptive automation with assessments of operator cognitive state. Figure 1 illustrates how the Honeywell team approached the research and development along the two themes via a spiral development. Through the phases, cognitive state classification has been matured from a laboratory, stationary setup with 32 EEG leads to a mobile system with six leads. Multiple mitigations were explored throughout the phases, covering the gamut from task offloading, task sharing, and task/information scheduling to modality management. Each experiment detailed in this report built upon what was learned in previous experiments.

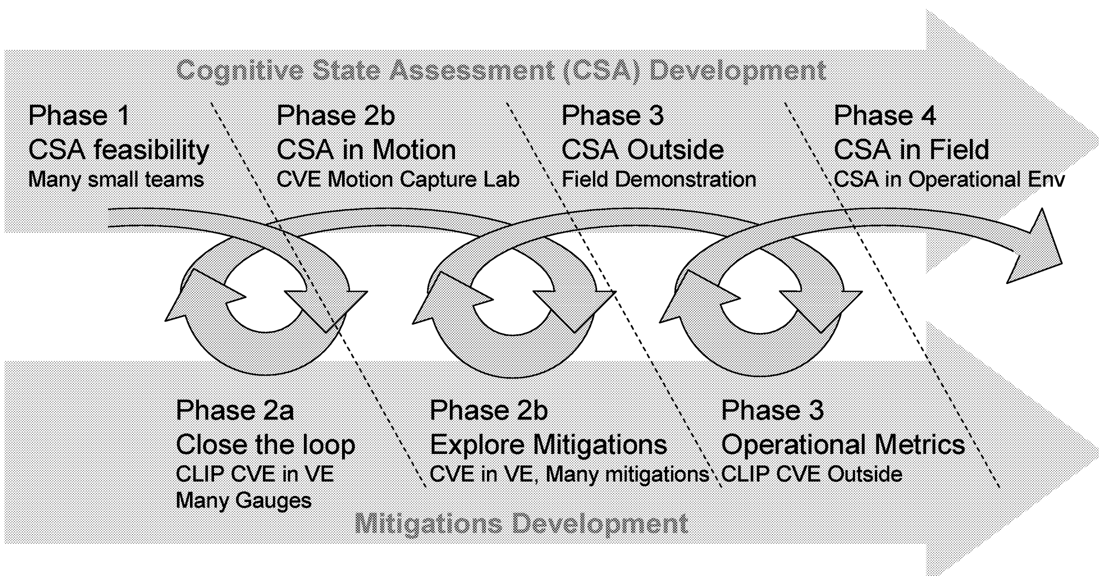


Figure 1. Spiral development of two parallel research thrusts.

The Concept Validation Experiment (CVE) of Phase 2a concentrated on integrating the cognitive state assessment technologies of Phase 1 into a closed-loop integrated prototype (CLIP). Honeywell concentrated on developing a CLIP that triggered (based on context and cognitive state assessments) information management mitigations to reduce demands on attention and improve participant performance. Cognitive state assessment was driven by a comprehensive suite of sensors, including a 32-lead EEG system, pupilometry, electrodermal response (EDR), ECG, and electromyogram (EMG). These sensors served as inputs to five cognitive state gauges (Arousal Meter, Stress Gauge, Engagement Index, eXecutive Load Index (XLI), and P300-driven novelty detector), as well as more straightforward measures of physiology such as heart rate and Interbeat Interval (IBI). The participants' interaction with the CLIP was evaluated in a virtual environment (VE)

with tasks that approximated the cognitive load of military operational tasks (identification of friend or foe, engagement of foes, navigation, and communications).

In Phase 2b, two CVEs were conducted. Carnegie Mellon University (CMU) performed a CVE that took cognitive state classification one step further by focusing on the ability to detect cognitive state in a (semi-) mobile virtual environment. In parallel, the Institute for Human and Machine Cognition (IHMC) performed a CVE that studied a range of mitigations, utilizing the rapid prototyping afforded by evaluating the CLIP in a VE. Specifically, the IHMC CVE focused on the development of four different mitigation strategies with military-relevant tasks (navigate to an objective, engage foes, attend to radio communications, coordinate medevac, and target detection). Both virtual environments were chosen because of the flexibility they offered in creating scenarios that were operationally realistic. These environments also provided the ability to manipulate the attentional demands associated with tasks. Situating tasks within these virtual environments allowed experimenters to precisely relate simulation events to neurophysiological states assessed by the gauges. The two virtual environments also provided insight into the performance of the gauges under different levels of mobility.

Phase 3 conducted three evaluations. A field evaluation demonstrated, for the first time, an EEG in a completely untethered, mobile setting. Based on what was learned, a fully mobile CVE was conducted outdoors, outside the laboratory, to evaluate CLIP mitigations driven by cognitive state assessment. Specifically, the CVE was conducted to investigate the efficacy of two mitigation strategies outside the laboratory in a wooded field environment: a communications scheduler and a tactile cueing mitigation. The communications scheduler mitigation was driven by an assessment of the participant's current cognitive capacity to process incoming information, in order to improve decision making under high task load conditions. A tactile cueing mitigation was created to support the participant's navigation along a complex route. The evaluation was conducted to demonstrate whether cognitive capacity to perform under differing task loads could be detected using neurophysiological sensors and then if the adaptive automation/mitigation would appropriately regulate information flow.

Finally in Phase 3, the Honeywell team worked with the Army at the Aberdeen Test Center (ATC) to demonstrate the cognitive state assessment of a Joint Task Force (JTF) Commander performing in an operational exercise - the Joint Distributed Freeplay Event (JDFE) at Mulberry Point MOUT site. Using the variations in the cognitive workload required of the scenario, the Honeywell AugCog team evaluated the classification techniques previously used in the laboratory and field tests to classify performance. EEG data collected in the context of the JDFE event provided an opportunity to assess the potential for Honeywell's real-time classification approach in an operationally relevant task environment. EEG data was collected from three different commanders and submitted to the refined classification approach.

Phase 4 focused on operational feasibility. The culminating test event to the IWIIUS/AugCog program was an evaluation in a MOUT environment at ATC performed over a two-week training period using a platoon of Soldiers. The overall objective was to assess Soldier workload levels during various operational tasks requiring different levels of cognitive and physical engagement and demonstrate the effectiveness of the AugCog techniques as measures of cognitive loading during mission phases on key leadership

positions. The team evaluated the effectiveness of sensor-driven assessment of cognitive state, looking at both physiological (ECG) and neurophysiologically (EEG) based sensors. For the first time in the program, the Honeywell AugCog team explored a sensor fusion approach to cognitive state classification.

The remainder of this report describes in detail the experiments outlined above. Chapter 2 describes the Augmented Cognition CLIP architecture in general terms. Each phase is described in detail in Chapters 3, 4, 5, and 6, where the specific instantiations of the CLIP for that phase are described in detail in each chapter.

2 Closed-Loop Integrated Prototype

Throughout the phases of the Augmented Cognition (AugCog) program, Honeywell adhered to the same basic system architecture. The Honeywell closed-loop integrated prototype (CLIP) is depicted in Figure 2.

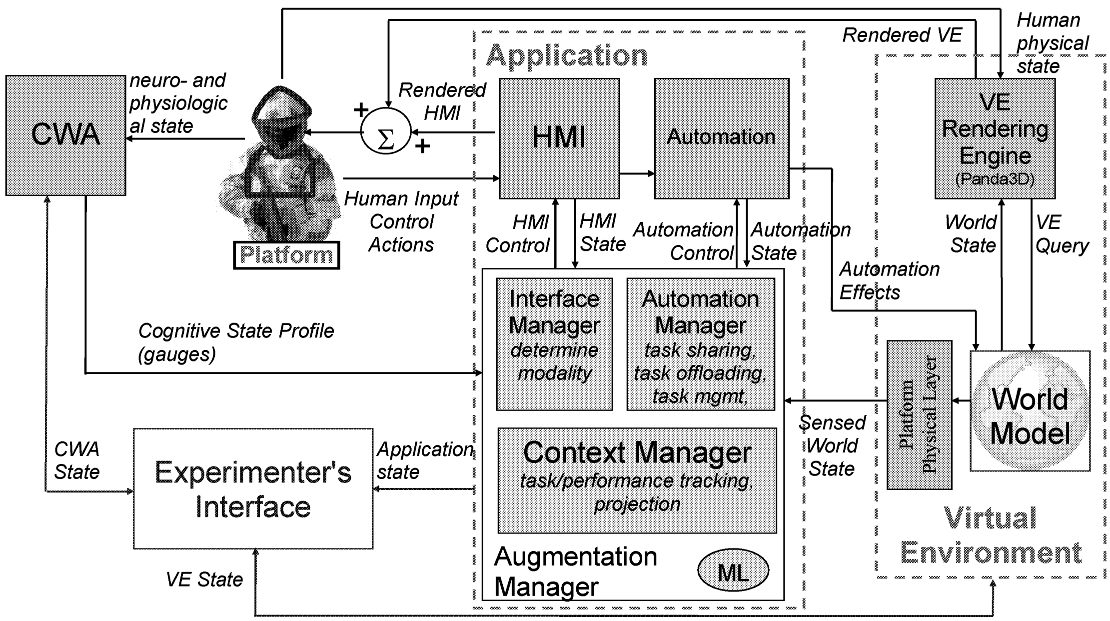


Figure 2. CLIP demonstration architecture.

The architecture is made up of the following components:

- **Cognitive Workload Assessor (CWA)** combined all the psychophysical measures of cognitive state available to the system to produce a single, extensible cognitive state profile (CSP) containing multiple dimensions of cognitive state.
- **Application** is the domain-specific portion of the AugCog system. As such, it contained the:
    - **Human-Machine Interface (HMI)**, where the human interacts with the system.
    - **Automation**, where tasks could be partially or wholly automated.
    - **Augmentation Manager (AM)**, where decisions are made on how to adapt the work environment to optimize joint human-automation cognitive abilities for specific domain tasks. The AM was composed of three components:
        - **Interface Manager**, responsible for realizing a dynamic interaction design in the HMI.

        - **Automation Manager**, responsible for the level and type of automation, and when it is applied.
        - **Context Manager**, responsible for tracking tasks, goals, and performance

- **Virtual Environment** (VE) was a simulated approximation of the real world. In Phases 2a and 2b, the CLIP was tested in a VE. The simulated environment consisted of three components:
    - **World Model**, which modeled all aspects of the world that are of interest to the simulation,
    - **VE Rendering Engine**, which generated a pictorial view into the World Model, and
    - **Physical Platform Layer**, which interacts with the World Model by "sensing" the "state" of the outside world and by impacting the "state" of the outside world.

Experimenter's Interface was used with the AugCog system to drive experiments and to give the experimenter both some insight into the workings of the system and control over some events within the system.

3 Augmented Cognition Program Phase 2a

3.1 Phase 2a Introduction

3.1.1 Phase 2a Research Team

The Honeywell Augmented Cognition (AugCog) team in Phase 2a consisted of the collaborative efforts of Honeywell Laboratories, Carnegie Mellon University (CMU), Clemson University, Columbia University, City College of New York, Human Bionics, Institute of Human and Machine Cognition (IHMC), Oregon Health and Sciences University, UFI, University of New Mexico, and University of Virginia. In addition, the team was advised by the Natick Soldier Research, Development and Engineering Center (NSRDEC). Phase 2a of the program encompassed work done between June 1, 2003, and December 31, 2003.

3.1.2 Phase 2a Research Objectives

The objective of Phase 2a was to develop technology that manipulated cognitive state. The focus of this phase was to conduct feasibility testing of closed-loop systems in cognitive environments. Overall, the metrics of Phase 2 (both Phase 2a and Phase 2b) were to detect cognitive state shifts in less than two seconds and trigger a cognitive state manipulation (mitigation) within one minute of the onset of the cognitive state shift.

3.1.3 Phase 2a Development Plan

In Phase 2a, Honeywell concentrated on developing a closed-loop system that triggered information management mitigations to reduce cognitive overload and increase participant performance. Existing cognitive gauges, as well as new gauge development, were integrated into the closed loop integrated prototype (CLIP). At this stage of the program, the CLIP was evaluated in a virtual environment (VE) to approximate the cognitive load of operational tasks.

3.2 Phase 2a Attention Bottleneck

The Honeywell team focused primarily on the cognitive bottleneck of attention. Many varieties of attention were considered to optimize their distribution: executive attention, divided attention, focused attention (both selective visual attention and selective auditory attention), and sustained attention. Breakdowns in attention lead to multiple problems: failure to notice an event in environment, failure to distribute attention across a space, failure to switch attention to highest priority information, or failure to monitor events over a sustained period of time. The Attention Bottleneck was important to the Future Force Warrior (FFW) program because it directly affected two cornerstone technology thrusts within the FFW program: Netted Communications and Collaborative Situation Awareness. Netted Communications will afford unparalleled knowledge and information exchange. Situation awareness is necessary for the individual, and Collaborative Situation Awareness is critical for the unit. Thus, the appropriate allocation of attention is the cornerstone of achieving situation awareness and mitigating information overload.

3.3 *Phase 2a System Design and Architecture*

3.3.1 Initial CLIP Overview

The Honeywell AugCog system was developed under the spiral development process, where development was iterated, and each iteration was a complete cycle of the requirements, design, implement, and test phases. The initial CLIP implementation is depicted in Figure 3.

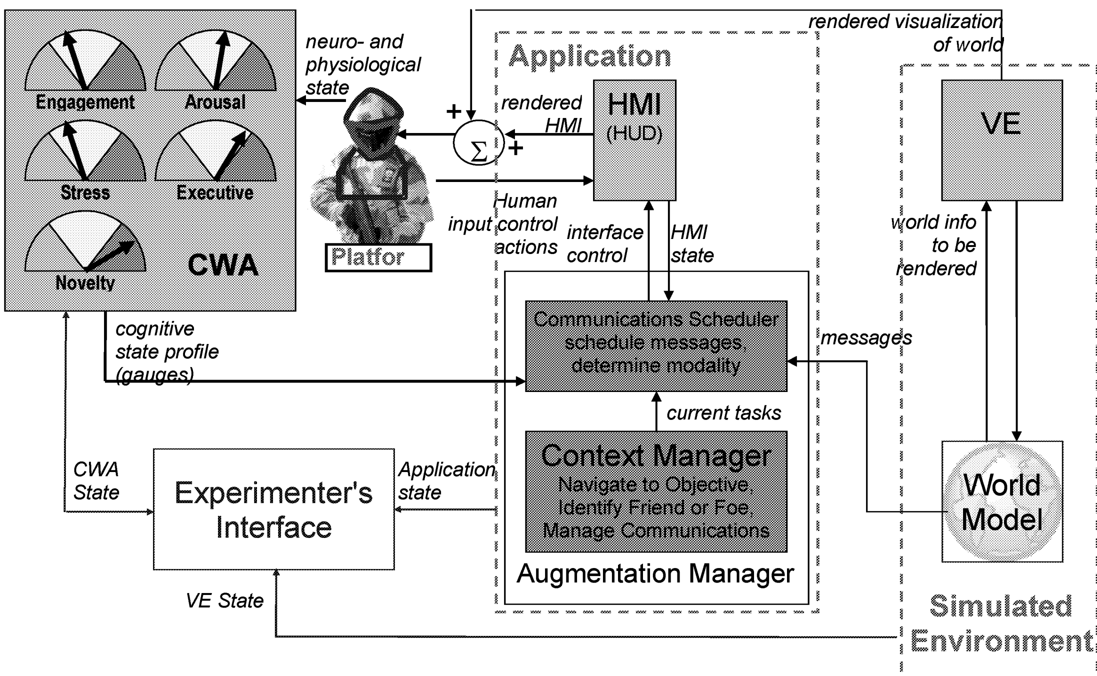


Figure 3. Initial CLIP implementation.

The Augmentation Manager (AM) reasoned about two primary tasks identifying friend or foe (IFF) and navigating the environment toward an objective, and one secondary task (attending to incoming communications). The principal mitigation strategy was the Communications Scheduler, which had the ability to prioritize, schedule, and modify incoming messages and present them to dismounted Soldiers in a manner that improves their performance on the primary tasks in the presence of demands placed by the secondary task. The Communications Scheduler took as input the CSP of the Cognitive Workload Assessor (CWA), the priority and participant of the incoming messages, and the state of the primary tasks in order to schedule the incoming messages. For more details on the Phase 2a CLIP configuration, see Whitlow et al., 2004.

3.3.2 CWA Gauges

This section describes Phase 2b sensor suite development for the Honeywell AugCog system. The Honeywell AugCog team developed an integrated, comprehensive suite of sensors, including electroencephalogram (EEG), pupilometry, electrodermal response (EDR), electrocardiogram (ECG), and electromyogram (EMG). For the Concept Validation Experiment (CVE), five gauges (Arousal Meter, Stress Gauge, Engagement Index, eXecutive Load Index (XLI), and P300-driven novelty detector) were created or

modified from previous versions to establish the cognitive state profile (CSP) of the participant.

The subsequent sections describe the gauges. For each gauge, the following information is detailed:

- Measures used as input (e.g., electroencephalogram, electrocardiogram
- Additional information about measures (e.g., interbeat interval (IBI), signal trial, pupil diameter, etc.)
- Processing done, methodology used, and advantages of the approach
- Cognitive states measured (e.g., comprehension, level of engagement, etc.)
- Levels measured (low, medium, high)

3.3.2.1 Engagement Gauge

The Engagement Index, as described by Freeman, Mikulka, Prinzel, and Scerbo (1999), was a measurement of how cognitively engaged a person was in a task or a person's level of alertness. Adaptive systems have used this index to drive control of the automation between manual and automatic modes. In fact, the index has been used to successfully control an automation system for tracking performance and vigilance tasks (Freeman, Mikulka, Prinzel, & Scerbo, 1999; Pope, Bogart, & Bartolome, 1995; Mikulka, Scerbo, & Freeman 2002).

To first replicate the work of Freeman et al., electroencephalogram was recorded from sites Cz, Pz, P3, and P4 with a ground site midway between Fpz and Fz. A running average of powers for different frequency bands was obtained using the following electroencephalogram frequency bands: alpha (8-13 Hz), beta (13-22 Hz), and theta (4-8 Hz). From these moving averages, the Engagement Index (beta/ (alpha + theta)) was calculated at regular intervals. Prinzel, Hadley, Freeman, and Mikulka (1999) reported that adaptive task allocation may be best reserved for the endpoints of the task engagement continuum; therefore, two levels of engagement were measured (low, high). The Engagement Index reflected the selection and focus on some aspect or task at the expense of the other competing demands—a measure of focused attention. High levels of engagement reflected selection and attentional focus, whereas lower levels of engagement indicated that the participant was not actively engaged with some aspect of the environment.

3.3.2.2 Stress Gauge

The IHMC developed a composite *Stress Gauge*, which measured physiologic changes in electromyogram, electrocardiogram, electrodermal response, and pupil diameter (root mean square (RMS) to detect the participant's response to changes in cognitive load within the virtual environment. The gauge used a weighted average of the four inputs (or any combination of a subset of the four sources) to indicate a normalized stress level. The gauge was used to detect cognitive stress related to managing multiple competing tasks on a moment-to-moment basis.

3.3.2.3 Arousal Gauge

Clemson University's *Arousal Meter* derived autonomic arousal from the cardiac IBI. Heart rate varies over time in response to moment-to-moment task demands, and these variations correlate with autonomic nervous system activity. The aim was to determine if autonomic arousal changes reflected a participant's response to a dynamic threat environment. IBI, the time between R-spikes in the electrocardiogram, or the time between heartbeats (see Figure 4) was derived from the electrocardiogram at 1-millisecond accuracy.

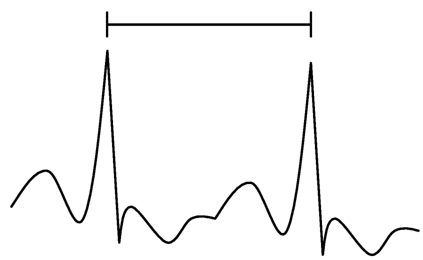


Figure 4. Interbeat interval.

A three-lead electrocardiogram was used to detect R-spikes and derive millisecond resolution IBIs that are then resampled at 4 Hz. Interbeat intervals of 16, 32, or 64 seconds were stored, and then a Fast-Fourier transform (FFT) was computed on the data. A sliding window was established such that every 0.25 second a new FFT was computed. When the FFT was computed, the high-frequency peak was identified, and the power at that peak, termed respiratory sinus arrhythmia (RSA), was stored. Once 1 minute's worth of FFT results were stored, the Arousal Meter began to generate a standardized arousal, computed every 0.25 second using a z-score standardization and the running mean and standard deviation of the RSA values. A standardized "arousal" score was derived [$-(x-\mu/\sigma)$], which drove the Arousal Meter. The gauge had three levels (low, medium, and high) to measure arousal. Increases in this score were associated with increased autonomic arousal and decreases with decreased autonomic arousal. A state shift was operationally defined as a score that changes from negative to positive. The Arousal Meter had approximately a 1-second resolution, but performed analyses four times a second for redundancy. The advantage to the approach was the computational efficiency, which resulted in a process that computed in real time on a low-power processor. In addition, since ECG is a strong signal, the data acquisition process was robust to participant movement.

3.3.2.4 Executive Load Gauge

Human Bionics developed the XLI. It operated by measuring power in the electroencephalogram at frontal (FCZ) and central midline (CPZ) sites. The algorithm used a weighted ratio of delta + theta/alpha bands calculated during a moving two-second window. The current reading was compared to the previous 20-second running average to determine if the executive load was increasing, decreasing, or staying the same. The index was designed to measure real-time changes in cognitive load related to the

processing of messages. This gauge was previously validated to discern trial difficulty in a continuous performance high-order cognitive task battery.

3.3.2.5 P300 Novelty Detector Gauge

Columbia University and the City College of New York created the *P300-driven Novelty Detector*, which measured a person's attention to a salient, task-relevant auditory probe that consistently preceded the arrival of an important communication. Unlike the other gauges, the P300 reflected a specific event-related response that assessed whether or not the participant attended an auditory probe. The mitigation premise was that if the participants did not attend to the probe due to lack of appropriate attentional resources, they could not process the incoming message. The gauge included frontal and parietal electrodes (as many as were feasible, since more electrodes provided more robust signals). The primary phenomenon measured was availability of attentional resources. Of particular interest was the P300 response within a task environment much richer than was typically the case in an experiment setting.

3.3.3 Communications Scheduler

For the Phase 2a CVE, Honeywell mitigated the attention bottleneck via information management strategies to manage incoming and outgoing netted communications. The Communications Scheduler scheduled and presented messages to the Soldier based on the CSP, message characteristics, and current context (tasks). The Communications Scheduler incorporated knowledge on the current context (i.e., tasks) of the Soldier and the current cognitive state via the cognitive state profile. In addition, the Communications Scheduler reasoned over communications and its associated priority, category, time requirements, response requirements, associated task, interaction requirements, source, and status. Based on these inputs, the Communications Scheduler could pass through messages immediately, defer and schedule non-relevant or lower priority messages, escalate higher priority messages that were not attended to, divert attention to incoming higher priority messages, change the modality of message presentation, or delete expired or obsolete messages.

3.3.3.1 Mitigation Strategies

There are five broad categories of possible mitigations in an AugCog system:

- Information Management
- Modality Management
- Task Management
- Task Offloading
- Task Sharing

The Communications Scheduler concentrated primarily on information management, although its ability to change audio messages to text was a form of modality management as well.

3.3.3.2 Message Priorities

All messages had a priority associated with them, depending on how critical they were. There were three priorities with the following definitions:

- High Priority: mission-critical and time-critical
- Medium Priority: mission-critical only
- Low Priority: not critical

At times when the augmentation was in effect, messages were scheduled according to certain rules. High-priority messages were mission-critical and time-critical, which means they must be heard and understood as soon as they arrive. Medium-priority messages were mission-critical but had a larger time window to work with. A medium-priority message could potentially be deferred if the system found that the Soldier was highly engaged in another task. All medium-priority messages were played before the end of the mission. Low-priority messages were not mission-critical or time-critical. They were presented if the participant was not engaged in another task. If the system found that the participant was engaged in another task, the low-priority message was presented in text format in a message window.

3.3.3.3 Message Tones

High-priority messages had a tone played before they were presented. If the system found, based on the cognitive state assessment, that the participant was highly engaged in a task, it played the same tone before the message, but louder and more salient. If the system found that the participant missed a high-priority message after it had been presented, it repeated the message once using the same tone, but louder again. In summary, three versions of the same tones were associated with high-priority messages.

Medium-priority messages also had a tone played before they were presented. It was recognizably different than the high-priority tone. Medium-priority messages were also repeated once if the system found that the participant missed a message, but the tone remained the same. It did not change in loudness like the high-priority tone did.

Low-priority messages did not have a tone associated with them. Low-priority messages were not repeated.

3.3.3.4 Message Window

For this study, only low-priority messages appeared in the message window, located at the bottom of the Soldier's field of view, as illustrated in Figure 5. Low-priority messages were indicated by a blue square.

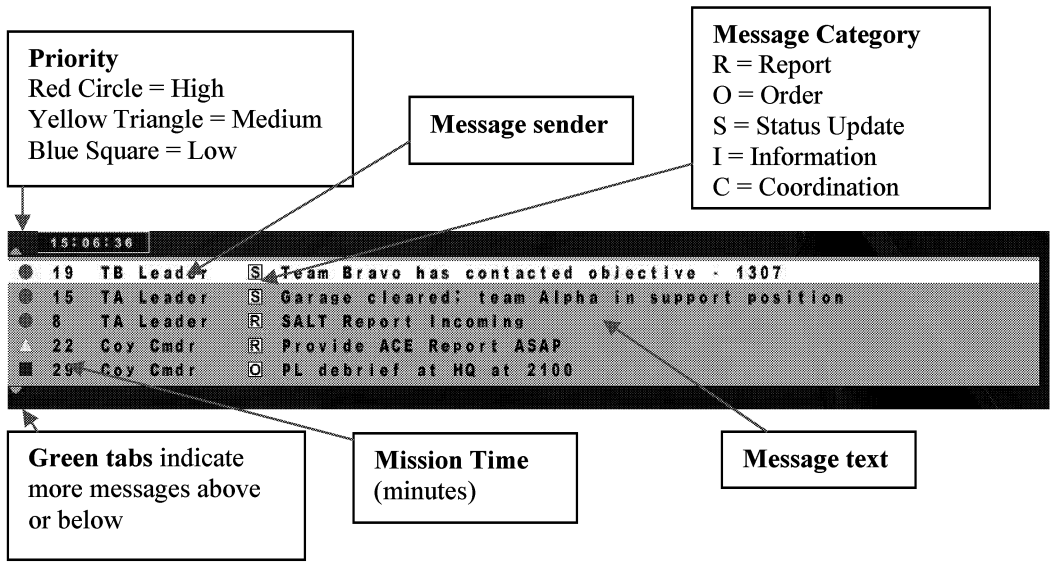


Figure 5. Message window.

3.3.4 Virtual Environment

3.3.4.1 Requirements

The Honeywell Augmented Cognition program required a VE with a representative Computer Generated Force (CGF) with appropriate fidelity to support sensor-suite validation and concept validation of the first closed-loop integrated prototype CLIP. Additionally, the CGF needed to provide a realistic, tactically correct MOUT battlefield environment. The VE needed to provide opposing forces (OPFOR) and friendly (blue) forces (BLUFOR) that would be controlled either by “botAI” (automated behavior scripts) or additional human operators. The VE needed to be of sufficient fidelity to represent the visual complexity of a MOUT environment in order to appropriately tax a participant’s workload. The fidelity of the Phase 2a CVE required that the VE have the following properties to add to the realism and immersiveness of the environment:

- Visually complex MOUT world
- Building interactivity to allow participant to enter buildings
- Three-dimensional world for mobility in the lateral and vertical directions
- Several participant behavior characteristics, including crouching, running, walking, jumping, firing weapon, climbing stairs, and depreciated health upon sustaining enemy attack
- Team members (BLUFOR) with the following characteristics: ability to fire at enemy, defend a position (or objective), move realistically, follow the participant, navigate to an objective, be tasked by the participant, and depreciate in health upon sustaining an enemy hit
- Audio to provide environmental sounds, weapons, and ability to insert audio messages from external Communications Scheduler

3.3.4.2 Current Implementation of the VE

Honeywell developed a desktop-PC VE based on a modified Quake3 TeamArena game engine. This environment provided high-fidelity dismounted combat operations that included tasking autonomous subordinates. The Quake3 game engine had the required fidelity for initial experimental validation.

The VE, illustrated in Figure 6, depicted a small area of a city, with realistic textures and detailed models, but with limited interactivity (most doors do not open or close, crates do not move, etc.). The city was composed of narrow streets surrounded by two- and three-story buildings. The environment had an industrial appearance. The participant entered into the environment in one of many predetermined locations in the map. In addition to the participant, there were some number of simulated players (bots), some OPFOR, and some blue forces BLUFOR. These forces were presented both at street level and above as snipers. The specific numbers of OPFOR and BLUFOR were adjusted at runtime. Each bot was assigned a skill level between 1 (easy) and 5 (hard). Therefore, workload was easily adjusted by manipulating the task load. Each player (participant or bot) had a realistic visual representation, with subtle details (primarily color and pattern of uniform) distinguishing BLUFOR from OPFOR.

The participant performed tasks in the environment using a combination of keyboard and mouse controls. The controls allow the participant to look around the world and to move (walking forward or backward, sidestepping left or right, jumping, and crouching).

Figure 6. Honeywell FFW virtual environment.

3.4 Phase 2a Concept Validation Experiment (CVE)

The CVE study was the first planned evaluation of the Honeywell CLIP in an FFW-relevant environment. Before a new gauge was integrated into the CLIP, each performer validated his/her performance in the controlled laboratory outside the full Honeywell sensor suite. The CVE offered the performers an opportunity to integrate the previously validated gauges into a more complex environment, and it offered Honeywell the opportunity to validate the gauge's effectiveness with an FFW-relevant task.

3.4.1 Experiment Objectives

The main thrust of this experiment was to identify the performance advantages and/or disadvantages associated with neurophysiological- and physiological-based gauges for determining cognitive state of the dismounted Soldier in the MOUT environment and moderating the high-workload, attentionally demanding states through automated mitigation strategies.

The goal of this research program was to develop an intelligent methodology for augmenting human cognition in response to changes in operator cognitive state. In response to increased cognitive stress, workload, working memory demands, attention tunneling, etc., such a system intelligently scheduled incoming information flow via sensory displays and triggered automatic systems to improve performance on a given complex task.

3.4.2 Experiment Hypotheses

The experimental hypotheses were focused on how to assess the performance and workload effects for completing the primary task of navigation and the secondary tasks of friend or foe identification and receiving and processing communications. The experiment evaluated the effectiveness of the AugCog mitigation strategy on the participants' overall performance. Overall performance was measured through collection of several metrics. The effectiveness of the mitigation strategy was determined by the response to messages and general situation awareness metrics. Specifically, the hypotheses were:

1. The optimal scheduling of information enhanced the Soldiers' performance on the communications management task, while not seriously degrading their performance on the other tasks of Navigate to Objective and IFF.

2. When the participants' tasks were augmented with the mitigation strategy for communications scheduling, their performance in message response and situation awareness was enhanced.

3. Performance enhancements might have included faster response times to messages, better message comprehension, greater overall SA, and lower perceived levels of workload.

4. Increases in workload (more and smarter enemy combatants) would reduce performance. The effectiveness of the mitigation strategy may only be present in the higher workload levels.

3.4.3 Operational Scenario

The CVE participant played the role of a Soldier navigating through an urban environment toward an objective. Soldiers identified friend or foe as they proceeded toward their objective. Periodically, they received incoming communications from their subordinates as well as higher echelons. They received status updates, mission updates, requests for information, and reports; these incoming communications were a primary source of their situation awareness.

3.4.3.1 Navigate to Objective Task

The participant was a platoon leader (PL) leading his or her platoon (1 Platoon) through a hostile urban environment to the objective. The objective is a room that the participant is meant to enter and clear to end the mission. Clearing a room in this scenario means neutralizing any enemy present in the room that is the final destination. When the

participant reaches the objective and is finished clearing the room, the participant is to report verbally to the company commander that he or she has reached the objective and is in place.

The participant in the demonstration followed a predetermined path through a small area of a city. The visual complexity of the environment contributed to the participants' workload. Participants were given a top-down map (illustrated in Figure 7), and they practiced the route three times before the first instance that the trial required that route. There were four routes for 16 trials.

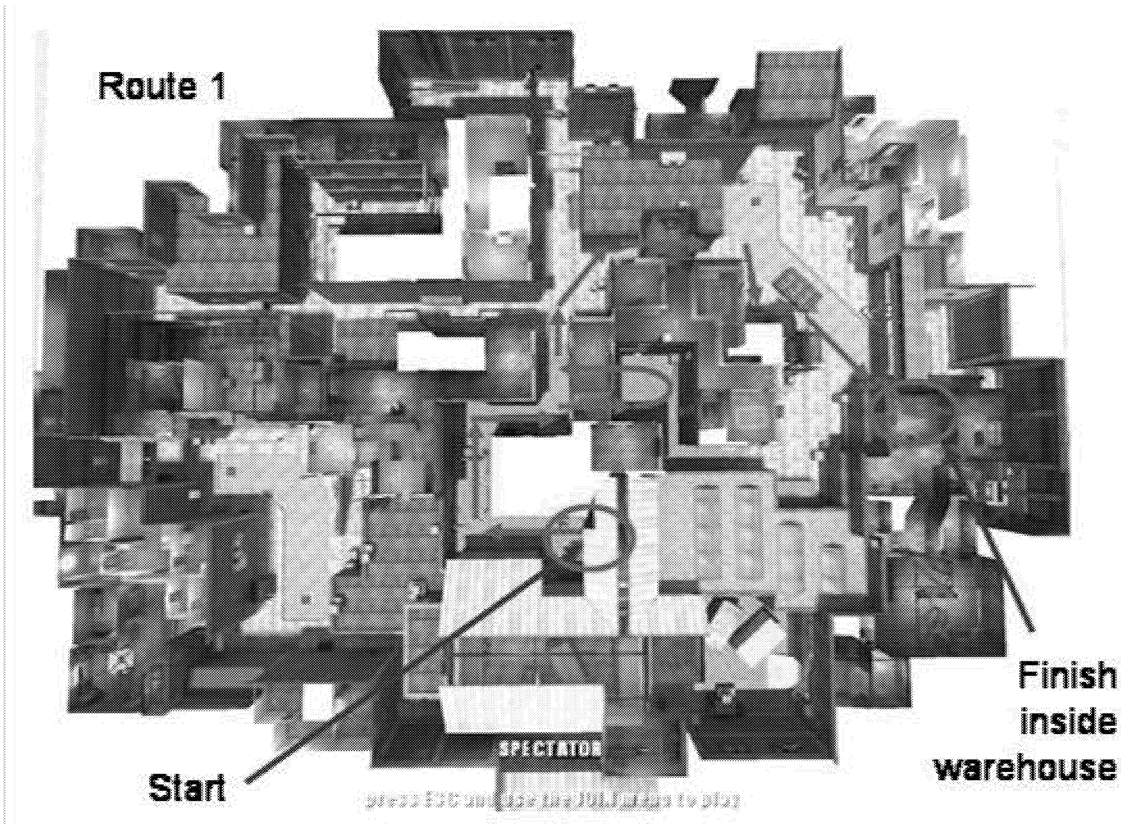


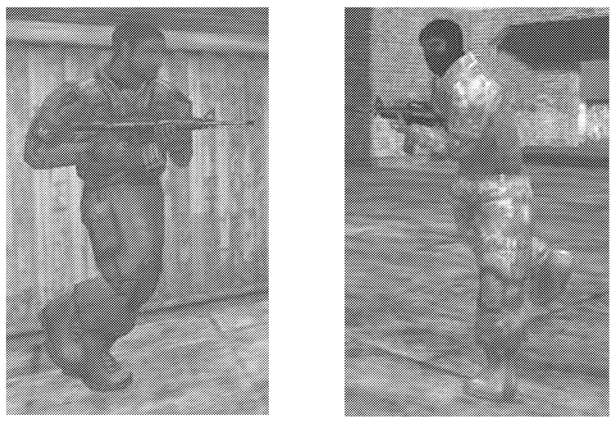
Figure 7. Route 1 for the navigate to objective task.

Instructions to the participant for the task were as follows:

1. Navigate to your objective as quickly as possible.
2. Avoid being shot.

3.4.3.2 Identifying Friend or Foe Task

One of the tasks in the VE was identifying friend or foe (IFF). Figure 8 shows both a BLUFOR Soldier and an OPFOR Soldier. The participant was faced with a specific number of enemy forces. These forces were presented both at street level and above as snipers. The enemy forces had logic for detecting the presence of the participant or other friendly forces and attacked with varying levels of success (depending on the workload and difficulty settings). The enemies were placed in a variety of locations.



Friend Foe

Figure 8. Friend and foe in the FFW virtual environment.

Instructions to the participant for the Identify Friend or Foe task were as follows:

1. Correctly identify and shoot the enemy. The enemy is wearing tan uniforms. The enemy will shoot at you.
2. Correctly identify and not shoot your team members. Your team members are wearing blue uniforms. Your team members will not shoot at you.

3.4.3.3 Communications Management Task

The platoon was divided into several teams that were taking different routes to the objective or were stationed in different locations near the objective. The platoon hierarchy and roles of the CVE participant are shown in Figure 9. The participant was the PL and had the dual role of leading Fire Team 1. The PL had three fire teams (one of which the participant was also leading), a support squad led by the platoon sergeant (PSG), and a security squad that was preventing more enemy troops from entering the participant's area of operations. The participant encountered enemies along the way, as well as members of his or her platoon.

The participant was in radio contact with his or her company commander, who was also in contact with the PLs for 2 Platoon and 3 Platoon and various other company commanders in the area. The company commander's job was to make sure that the participant's mission was being completed and to inform the participant of events occurring near the participant's area.

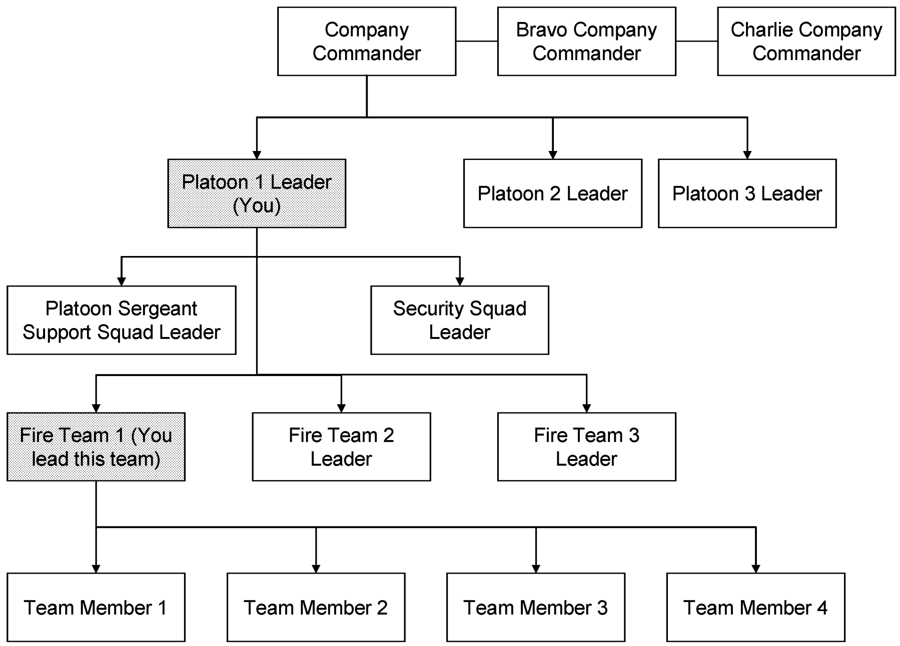


Figure 9. The platoon hierarchy.

As the PL, the participant's job was to maintain awareness of the progress of the mission by receiving and responding to messages from various platoon members, such as the fire team leaders, the PSG and team members. The participant also received messages from his or her company commander. Messages may or may not have been immediately critical to the mission; however, *all* messages were intended for the participant to hear. Some messages required the participant to respond, while others contained information being passed over the radio that did not require a response.

Instructions to the participant regarding the communications management task were as follows:

1. To respond appropriately to any messages requiring a response. Requests for a response may come in two forms:
    a. Some messages will ask a specific question. (e.g., *"PL, can you tell me how many snipers you have encountered?"*)
    b. Some will require you to acknowledge that you heard and understood the message. You should only acknowledge a message if you actually heard it AND understood it. (*e.g., "PL, we're sending troops into your area. Acknowledge.")*
2. If you understand a message that requires a response, but are unsure of the answer, it is important to answer nonetheless. In other words, there is no right or wrong answer, the important thing is to answer if you understood the question.

(e.g., *"PL, what is your current location?"* Answer: *"Current location unknown"* or *"halfway through my route"* or *"in an alley."*)

3. Suggest that the participant respond with short answers.

3.4.3.4 Manipulating Workload

One of the most difficult challenges was tailoring the scenarios to ensure that participants experience either high or low workload. A behavioral study (see Whitlow et al., 2004) revealed a number of dimensions where scenarios could vary in order to increase or decrease workload via task load manipulation. An example is the number of snipers. It was not sufficient to simply increase the number of snipers to increase moment-to-moment workload, especially if the participant would encounter the snipers one at a time. Increasing the number of snipers simply increased the number of times they were in a situation, rather than changing the workload of that situation. Thus, it was necessary to increase the number of snipers a participant was faced with simultaneously. Workload was observed to increase dramatically when participants were faced with multiple (well-placed) snipers in comparison to a single sniper.

The visual complexity of the VE contributed to the participant's workload. The participant was faced with a specific number of enemy forces. These forces were presented both at street level and above as snipers. The enemy forces had logic for detecting the presence of the participant or other friendly forces, and attacked with varying levels of success (depending on the workload and difficulty settings). The difficulty in each scenario in a block was the same (as defined by accuracy and intelligence of the opposing force); however, the start and endpoint and the route between them varied. This allowed repeated measures at a specific difficulty level but prevented participants from memorizing elements of the scenario.

3.4.4 Participants

Twelve males aged 18 to 33 (average was 24) participated in this study. All participants reported normal (N = 10) or corrected-to-normal (N = 2) vision and normal hearing, and none reported color vision deficiencies.

To minimize the impact that learning to play a first-person shooter game might have on the study and the data derived from the gauges, participants were chosen who had previous experience with first-person shooter games. The participants reported playing an average of 8.25 hours (range = 1-25 hours) of first-person shooter games per week. Two participants reported their skill level as "exceptionally good," five reported their skill level as "reasonably good," and four reported their skill level as "so-so."

The potential for experiencing simulator sickness was small, but participants were asked if they had experienced simulator sickness symptoms in the past. Only one participant reported experiencing symptoms previously. No one experienced simulator sickness during the CVE.

3.4.5 Experiment Design

3.4.5.1 Independent Variables

There were three independent variables: Workload level (low, high); AugCog mitigation strategy (present/absent), and route (4). The routes were constructed (and tested) to be functionally identical so that they could be used as repeated measures. Each participant completed four scenarios in each of the two workload conditions, in each of the AugCog conditions, for a total of 16 trials. The order of the AugCog mitigation condition was counterbalanced. Participants proceeded successively from the lowest to highest workload levels in the evaluation.

3.4.5.2 Experiment Design

The overall design was a within-subjects 2x2 design with the level of workload (2) and mitigation strategy (2) as within-subjects variables (see Figure 10).

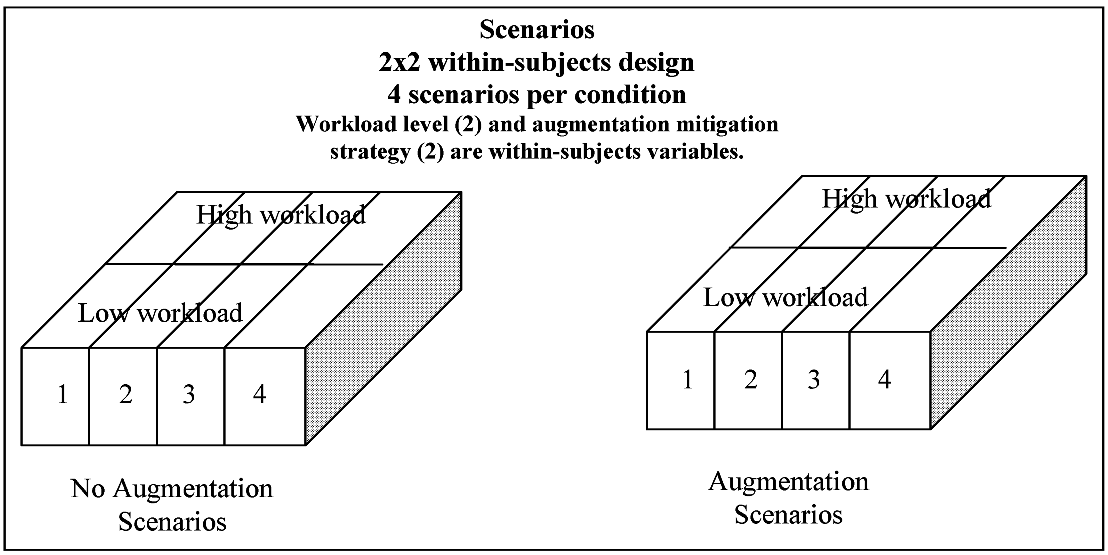


Figure 10. Experiment design for the CVE.

Each block of trials consisted of four scenarios (route) in which the participant navigated from Point A to Point B while identifying and engaging enemy targets and communicating with team members. The participants received each of the two workload blocks of scenarios in one of the mitigation strategy conditions before transitioning to the second condition. The presentation order of the mitigation strategy condition was counterbalanced between participants. The participant therefore conducted 16 trials, as illustrated in Table 1, where L = low, M = medium, H = high, N = no augmentation, and A = augmentation. Trial 17 was a repeat of trial 16, but the participant was standing and walking in place.

Table 1. Experiment trials.

Book A (8 Participants)																	
Trial	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Route	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	4
Workload	L	L	L	L	H	H	H	H	L	L	L	L	H	H	H	H	H
Mitigation	N	N	N	N	N	N	N	N	A	A	A	A	A	A	A	A	A
Book B (8 Participants)																	
Trial	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Route	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	4
Workload	L	L	L	L	H	H	H	H	L	L	L	L	H	H	H	H	L
Mitigation	A	A	A	A	A	A	A	A	N	N	N	N	N	N	N	N	N

The 12 participants were randomly assigned to one of the AugCog mitigation conditions seen first. The level of workload and mitigation strategies were a within-subjects condition that allowed each participant to receive all conditions.

3.4.6 Dependent Measures

Several categories of metrics were employed in this experiment. Objective performance measures included time to reach destination, distance traveled, destination accuracy, response time to message, number of opposing forces killed, number of times player is killed, number of times player shoots teammates, shooting accuracy, and reaction time to enemy combatants once they are in view. Gauge effectiveness metrics included: correlations of performance with each of five sensor gauges (Arousal Meter, Stress Gauge, Engagement Index, Executive Load Gauge, and P300 Novelty Detector). Workload measure included NASA (National Aeronautics and Space Administration) TLX (Task Load Index) Rating scales after each block of four trials and questionnaires. Situation awareness measures included probe questions regarding content of messages, enemy positions, and environment via a questionnaire at the end of each trial. Finally, preferences, acceptability, and qualitative feedback were gathered via a post-experiment questionnaire.

3.4.7 Experiment Protocol

The experiment protocol is summarized in Table 2. Participants were briefed about the experiment. The briefing presented all the information the participant needed to execute the mission, including descriptions of the goal (navigating to specific locations, securing locations), the friendly team (quantity, uniform appearance), the enemy forces (approximate quantity, uniform appearance, likely tactics), the route to be taken to the objective (possibly including a map), and a general description of the performance goals (reach the objective quickly, minimize friendly deaths, maximize enemies killed, attend to messages). After the sensing equipment was placed on the participant, the equipment was calibrated through a series of simple tasks. Participants were trained on the tasks they were to perform and familiarized with the VE. Each participant completed 16 trials, eight in each of the two workload conditions. The 17th trial repeats the scenario of trial 16, except the participant is walking in place. After the equipment was removed, participants were debriefed, and they filled out a questionnaire.

Table 2. CVE protocol.

Phase	Task	Time (min)
Briefing		20
	Introduction	
	Purpose of Assessment	
	Evaluation Personnel	
	Experiment Schedule	
	Consent Form	
	Demographics Form	
	NASA TLX Workload Scale Instructions	
Calibration & Familiarization		40
	Place sensing equipment on participant.	
	Calibration routines	
	Familiarize with virtual environment controls.	
	Practice sessions in virtual environment.	
	Experiment Scenario Instructions.	
	Put on helmet with eye tracker.	
Experiment Trials: Block 1 (Trials 1-4)		3
	Practice Route (before each trial)	
	Post-Trial Questionnaire (after each trial)	
	NASA TLX (after trial 4)	
Experiment Trials: Block 2 (Trials 5-8)		25
	Post-Trial Questionnaire (after each trial)	
	NASA TLX (after trial 8)	
Break		10
Experiment Trials: Block 3 (Trials 9-12)		25
	Post-Trial Questionnaire (after each trial)	
	NASA TLX (after trial 12)	
Experiment Trials: Block 4 (Trials 13-16)		25
	Post-Trial Questionnaire (after each trial)	
	NASA TLX (after trial 16)	
Experiment Trials: Trial 17		10
Post-experiment		30
	Post-Experiment Questionnaire	
	Debrief	
	Payment sheet	

3.4.8 Data Analysis Methodology

Data analysis routines were developed from sample datasets collected at the Pre-CVE (a dry run of the experiment). Data analysis fell into four primary categories:

1. *Sensor Data Quality Assessment* determined if the sensor data (e.g., electroencephalogram) was clean and free of artifacts that may compromise analysis.
2. *Gauge Assessment* determined if the gauges behaved in meaningful ways in the various experimental conditions.

3. *Performance and Workload Metric* assessed the changes in performance and metrics under various experimental conditions for the three tasks participants performed in the CVE.
4. *Mitigation Behavior Analysis* determined how the mitigation strategies behaved when driven by the CWA gauges.

3.4.8.1 Sensor Data Quality

The primary concern was to confirm that the CVE was getting reliable electroencephalogram signals since the hardware was relatively new, was an advanced research model (BioSemi), and was being deployed in proximity to other sensors such as an ISCAN eye tracker. The City College of New York (CCNY) sampled long trials for each participant as a preliminary signal quality analysis.

CCNY's primary concern was assessing if the collected electroencephalogram data was of high enough quality to support the detection of evoked responses and not necessarily for evaluating the signal quality for the other electroencephalogram-based gauges that are more derived and rely on average frequency bins such as Engagement Index and XLI. Honeywell and Human Bionics evaluated the quality of the gauge data and cleaned select samples with outlier analysis.

3.4.8.2 Gauge Assessment

Each gauge was validated during development to ensure that the outputs of the gauge could be interpreted in meaningful ways. Typically very basic, well-understood tasks were used to validate gauges at this stage; however, there is a large gap between simple laboratory tasks and tasks in the "real" world. Thus, once the gauges were integrated into the CLIP platform, they needed to be validated against FFW-relevant tasks to again ensure that the results of the gauge are meaningful. It was likely that some gauges were useful in conjunction with some tasks and not others. Thus, the final milestone of sensor integration was to empirically evaluate the current set of integrated gauges with FFW VE.

eXecutive Load Index (XLI): The operation of the XLI was validated across the two workload conditions and compared between the two augmentation conditions. First, the AugCog team correlated gauge output to workload conditions by first summing across all scenarios in each workload condition and comparing results with the low and high levels of CVE operation to identify significant mean cross-trial differences. Analysis of variance was used to compare results.

Arousal Meter: Heart rate varies over time in response to moment-to-moment task demand, and these variations correlate with autonomic nervous system activity. Specifically, autonomic arousal, measured with a derived cardiac IBI indicator called an "Arousal Meter," was evaluated across the two workload conditions in addition to more local changes driven by evolving task demands. The AugCog team determined if autonomic arousal changes reflect a participant's response to a dynamic threat environment. Specifically, the team assessed the autonomic response to the number and skill of hostile computer-generated forces (CGF)—represented by low and high workload.

Interbeat Interval(IBI): In addition to providing the raw measures for Clemson's Arousal Meter, the CVE also assessed how IBI (time in milliseconds between adjacent normal heartbeats) correlates with workload and augmentation trials. In addition to these trial-wide effects, IBI response was evaluated with regard to changing task loads within trials. It was expected that as task load increases from a single to three concurrent tasks, IBI would monotonically decrease, which indicates an increase in autonomic arousal.

Stress Gauge: It was anticipated that the composite Stress Gauge, which measured physiologic changes in electroencephalogram, electrocardiogram, electrodermal response, and pupil size, would detect the participant's response to changing cognitive load within the virtual environment. While the AugCog team anticipated seeing subtle, global differences between the two workload conditions, the team was more interested in how this gauge could detect cognitive stress from managing multiple competing tasks on a moment-to-moment basis. For example, it was expected that when participants encounter multiple hostile CGFs that are operating in close proximity to friendly CGFs, the composite Stress Gauge would reflect a systemic response to tracking multiple objects while discriminating friend from foe; moreover, this environment provides ample opportunities to induce cognitive stress from requiring participants to manage two primary tasks (navigation and IFF) with periodic communications management.

Engagement Index: The Engagement Index is an electroencephalogram power-based measure of moment-to-moment attentional engagement with the task environment. The specific formula (20 * beta (alpha + theta)) has been validated in several studies of attention and vigilance. Whereas the composite Stress Gauge reflects the cognitive load from managing multiple, competing demands, engagement reflects the selection and focus on some aspect at the expense of the other competing demands. For example, if the participants approached an intersection, they would likely attend to those cues that would help them make the correct navigational decision. Another likely scenario was when the participant approached some objective and received incoming sniper fire from an adjacent building. At this point, the participant narrowed the focus and engaged the building to locate the enemy sniper. High levels of engagement reflected selection and attentional focus, whereas lower levels of engagement indicated that the participant was not actively engaged with some aspect of the environment. There is a particular interest in how the Engagement Index and composite stress vary and complement each other as the task environment evolves; the AugCog team will explore this in correlation analyses between gauges within and across experiment trials.

P300 novelty detector: Also assessed was a participant's attention to a salient, task-relevant auditory probe that consistently preceded the arrival of an important communication. Unlike the other gauges, the P300 reflected a specific event-related response that assessed whether or not the participants attended an auditory probe. The mitigation premise was that if they did not attend the probe, they had not redirected attention and were not ready to receive and process an important communication.

3.4.8.3 Performance and Workload Metrics

The dependent measures were compared across low and high task load conditions. In addition, workload effects across the high task load and low task load were assessed for each of the performance metrics listed above. Subjective workload assessments via the

NASA TLX scales administered after every four trials were used to validate the workload induced by the high- and low-workload scenarios.

3.4.8.4 Mitigation Behavior Analysis

The mitigation strategies of the Communications Scheduler were driven by a set of rules that considered every possible combination of gauges. Three gauges (Engagement, Arousal, and Stress) were considered before a message was presented to determine the optimal method of presentation. Once a message was presented, two gauges (XLI and P300 novelty detector) were used to determine if the participant had the attentional resources available during the message presentation to attend to and comprehend the message. If these two gauges decided that the participant did not have sufficient resources, the message was repeated with a more salient tone to divert attention to the message.

The Mitigation Behavior Analysis focused on understanding how often various strategies were employed and under what conditions. Furthermore, the analysis looked into the gauge values used at the decision point of the Communications Scheduler to understand how the CWA gauges drove the resulting mitigation strategies.

3.5 Phase 2a CVE Results

This section details the findings from the CVE conducted by Honeywell at the IHMC in December 2004.

3.5.1 Sensor Data Quality

This analysis of the electroencephalogram data quality was based on the CVE data collected at IHMC using the 40-channel BioSemi system from visual calibration and two baseline runs on all participants.

The data for participants 8, 10, 11, 14, and 17 looked on the surface like clean electroencephalogram signals. However, only in participants 11 and 14 can one identify a trial-averaged evoked response 150 milliseconds following the tone marker. The signal was rather weak, so CCNY did not look at its spatial origin. The signal-to-noise ratio was not sufficient to detect it on a single trial basis. Participant 8 contained substantial artifacts that could be due to motion or electrostatic discharge. Participant 17 was reported to have timing errors. Participant 10 did not show any evoked response in the trial average.

Participants 4, 5, 6, 7, 9, 12, 13, 15, and 16 contained sections of data at random times and channels that looked like dynamic range overflow (when the signal exceeds the ranges –262142.96875, 262142.96875, the signal changes sign). This recording error made electroencephalogram analysis rather difficult, so data were not further analyzed for those participants. Hence, it must be concluded that this electroencephalogram data was not of sufficient quality for the purpose of P300 detection.

3.5.2 Gauge Assessment

This section summarizes the Gauge Assessment results. A more detailed discussion can be found Whitlow et al., 2004. High inter-participant variability led to generally non-

significant effects for the full 2 x 2 ANOVA (Analysis of Variance). One caveat for evaluating these gauges was that the CVE task environment, "first-person shooter" VE with incoming auditory messages, is fast-paced and required nearly constant allocation of substantial cognitive and perceptual resources. Thus it was unlikely to place participants in an underload condition; accordingly, the sensitivity of the gauge suite to underload could not be assessed.

3.5.2.1 Engagement Index

There was a numerical difference in the expected direction between all high-workload conditions (.256) compared with low workload (.196). There was also a numerical, and trending toward significant ($p < .14$), finding for augmentation with Augmentation ON (.082) compared with OFF (.370). Finally, there was also a numerical, though not significant, interaction between workload and augmentation manifested by a much greater difference between Augmentation ON and Augmentation OFF for high workload (delta is .183) compared with low workload (delta is .393).

It was encouraging to see a higher level of task engagement with Augmentation OFF as a possible indication that the CWA-driven smart Communications Scheduler was reducing attentional demands. Furthermore, it was expected, and was seen, that the greatest benefit, or difference, driven by augmentation under high-workload conditions. As an indicator of attentional load, Engagement Index values confirmed that participants had higher attentional requirements without the benefit of augmentation.

It was also interesting to note that the average for all treatments was above 0, which was the baseline established for each participant. This confirmed that the addition of auditory message management during the experiment trials increased the attentional requirements.

3.5.2.2 Clemson Arousal Meter

As was expected, the lowest arousal scores were seen during the first baseline, and the highest arousal scores were seen in a standing position. The physiological challenge of sitting relaxed versus standing while performing a task shows that the Arousal Meter was indeed functioning and responsive.

Hence it can be concluded from these data that performance on a computer task was significantly different from resting and performance on a computer task while standing was significantly different from resting. However, the cognitive Arousal Meter did not appear to be sensitive to detecting differences in cognitive loads influenced by performing computer tasks that vary in their task loads.

3.5.2.3 IBI

It was encouraging to find both significant trial-wide and within-trial effects. First, participants had significantly higher IBIs on augmented trials compared with non-augmented trials. This suggested that the augmentation intervention decreased participants' autonomic arousal, as indicated by a higher IBI. Effectively managing autonomic arousal has the potential for improving overall task performance by preventing Soldiers from migrating to the performance decrement area of the Yerkes Dodson curve.

Also encouraging was that IBI was sensitive to moment-to-moment changes in task load, as indicated by the significant effect of task load within trials. Furthermore, IBI reliably and significantly differentiated between high task load compared with both low and high task loads.

3.5.2.4 Human Bionic's XLI

The findings indicate that the XLI accurately differentiated between low-load and high-load conditions without augmentation in 10 out of 11 participants, while only participant 14 showed an opposite situation where the high condition indicated lower workload.

A 2 x 2 repeated-measures ANOVA was run to evaluate main effects: workload—high, low; augmentation—on, off. The analysis showed a marginally significant main effect of workload ($F = 3.276$, $p < .098$), indicating that the workload manipulation was reliably detected by the XLI gauge.

3.5.2.5 CCNY/Sarnof P300

Preliminary results from the responsiveness of the P300 gauge suggested that it detected an evoked response in real time more than 64% of the time (162 of 250 occurrences of high-priority auditory alerting tones across all participants for all trials). This result was encouraging, considering this was the first integration of a real-time P300 gauge in such a cognitively and perceptually rich task environment.

3.5.2.6 IHMC Composite Stress

Due to high inter-participant variability, the data within individual participants was analyzed to assess their response to the experimental manipulations. Ten of 12 participants were analyzed. Results indicate that six of ten participants showed significant effect for workload manipulation; furthermore, eight of ten participants analyzed also showed significant effect of augmentation. Accordingly, it was concluded that the composite Stress Gauge is responsive to the experimental manipulations. This was very encouraging, considering that the ISCAN eye tracker, which provided pupilometry as one of the four sub-indexes of stress, burned out after only four participants. In previous studies, pupilometry was the most salient contributor to predicting task loading.

3.5.3 Performance Analysis

3.5.3.1 Performance Results

Performance and workload effects were analyzed for completing the tasks of navigation, identification of friend or foe, and receiving and processing communications. Preliminary analysis on a subset of the data indicates an effect of workload on performance. Participants took more time to complete the navigation task, they were shot by the enemy more often, and their behavior was more evasive as compared with the low-workload trials. One of the main evaluations was to test the effectiveness of the AugCog mitigation strategy on overall performance, as measured by the communications management metrics of message response and situation awareness. Situation awareness measures were collected after completing the tasks. Good situation awareness (SA) is a cornerstone of an effective Warfighter; an effective mitigation strategy will enhance SA. Level of workload

appeared to mitigate the participants' SA as indicated by slightly better SA in the low-workload conditions as compared with high workload.

Since the team was augmenting secondary task performance (auditory communications management), the team evaluated whether this augmentation negatively affected primary task performance (navigating quickly and cautiously, IFF). Table 3 summarizes how the measures of performance map to tasks.

Table 3. Measures of performance.

Measure of Performance	Task	Augmentation Effect	Workload Effect
# times participant was hit by enemy fire	Navigation	No effect	Significant in expected direction
# times participant hit enemy	IFF	No effect	Significant in expected direction
Evasiveness (total time in view of enemy/(# times hits/)	Navigation	No effect	Significant in expected direction
Reaction time to enemy coming into view	IFF	Marginally significant ($p < .09$), decrement for Augmentation ON	No effect
Time to navigate to objective	Navigation	Significant ($p< .008$), decrement for Augmentation ON	Significant in expected direction
Hit rate (# times participant hits enemy/encounters with enemy)	IFF	No effect—significant WL x Aug interaction ($p < .047$)—better performance for Augmentation ON under high WL only	Significant in expected direction
Miss rate ((# times participant did not hit enemy/encounters with enemy))	IFF	No effect—significant WL x Aug interaction—better performance with Augmentation ON under high WL only	Significant in expected direction
Situation awareness (% correct on post-trial questions)	Messaging	No effect—significant WL x Aug interaction ($p < .020$)—better performance for Augmentation ON under high WL only	Significant in expected direction
Message acknowledgment (% correct response to within-trial auditory messages)	Messaging	Marginally significant effect ($p < .139$), improved performance with AugCog for both WL conditions	No effect

The current mitigation strategy produced a significant decrement in navigation time—participants were slower to reach their objective with Augmentation ON. This is likely due to the auditory tones inducing a strategy change whereby participants would pause when warned of an incoming high-priority message. Participants were also slower (marginally significant) in responding to enemies with Augmentation ON; one possible explanation for this is that the warning tone might have produced competition between IFF and messaging tasks. However, the augmentation strategy produced no decrement in primary task performance along two of three measures of performance for navigation and three of four measures for IFF.

There are also indications of some benefit from augmentation, especially in terms of minimizing the negative impact of high-workload conditions. This was supported by the significant interaction between workload and augmentation for situation awareness as well as hit rate. Under high workload, the Augmentation ON condition produced better situation awareness as well as higher hit rates on the enemy. This is consistent with many program findings that the benefit of augmentation is at the extremes of the workload continuum.

See Figure 11 and Figure 12 and for the 2 x 2 Analysis of Variance (ANOVA) results for each measure.

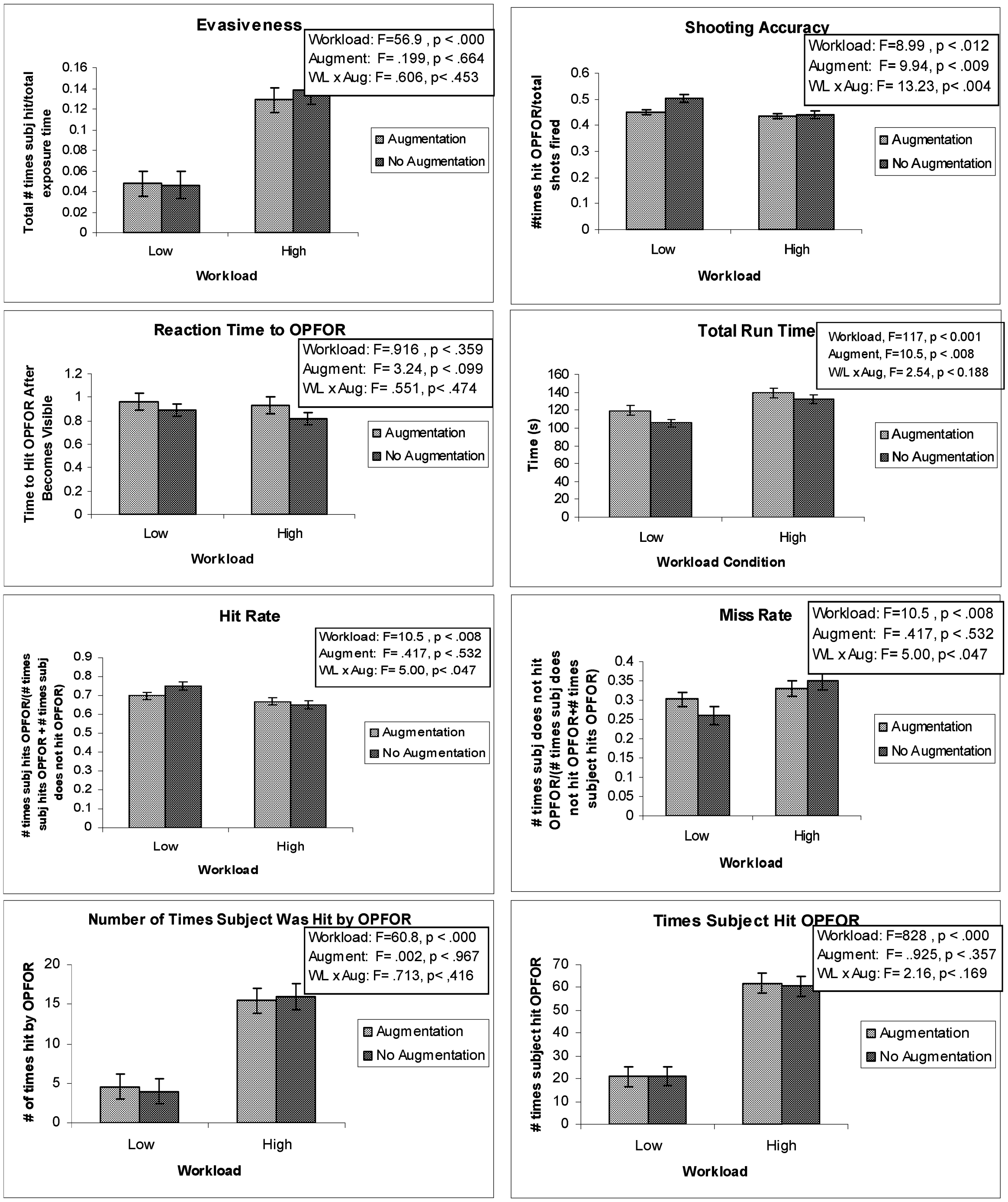


Figure 11. 2 x 2 ANOVA results for each measure.

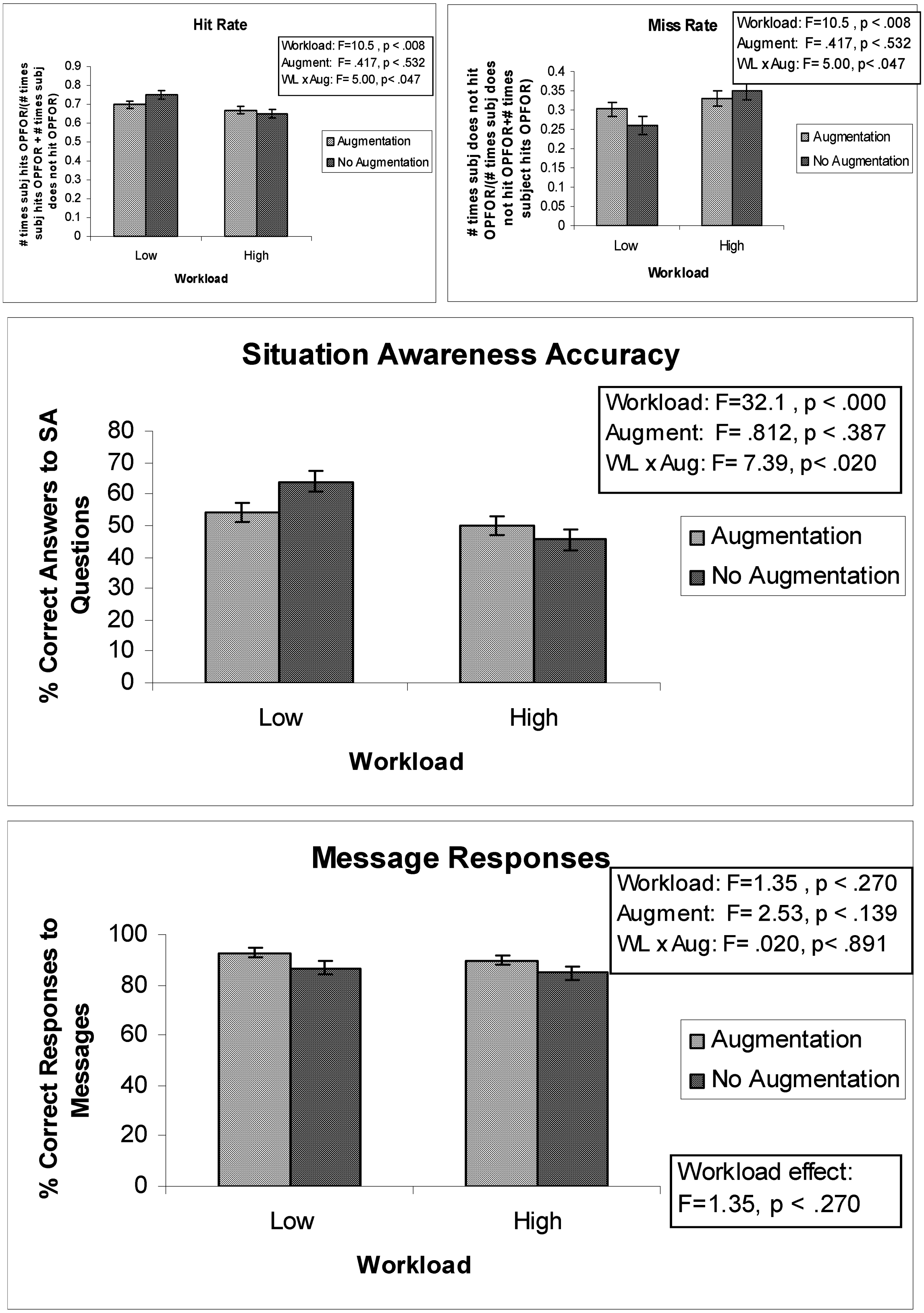


Figure 12. CVE metrics.

3.5.3.2 NASA TLX (Task Load Index) Results

TLX subscales consistently demonstrated significant effects of workload, except for Performance and Effort. See the figures below for the ANOVA results for each subscale.

Only Mental Demand and Overall TLX showed marginally significant effects of augmentation—in the direction indicating that augmentation actually increased mental demand. This is not entirely surprising, since one of the possible mitigation responses was to repeat messages, which potentially doubled the number of incoming auditory messages for some participants depending on their cognitive state during the trials.

See Figure 13 and Figure 14 for the 2 x 2 TLX results.

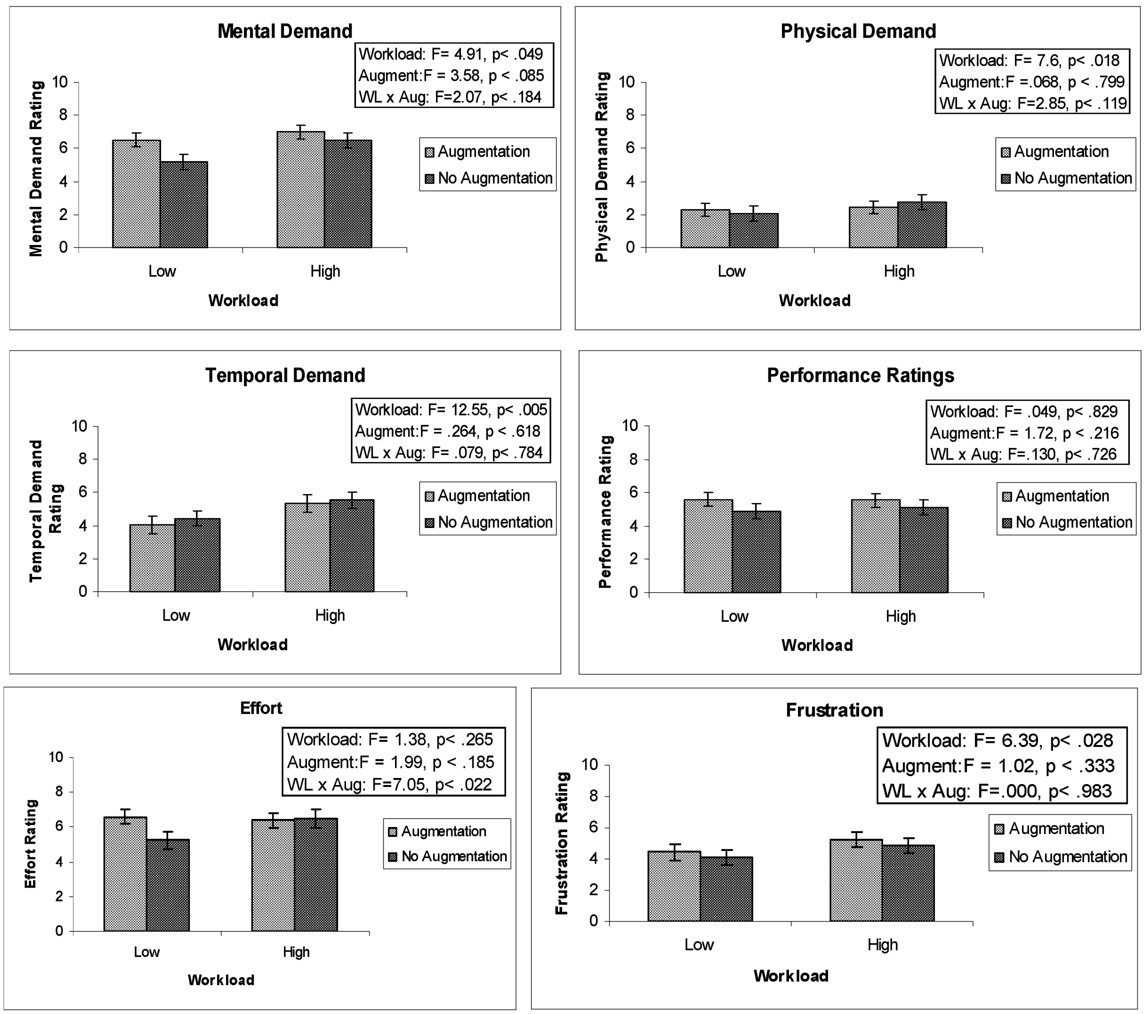


Figure 13. 2 x 2 TLX results.

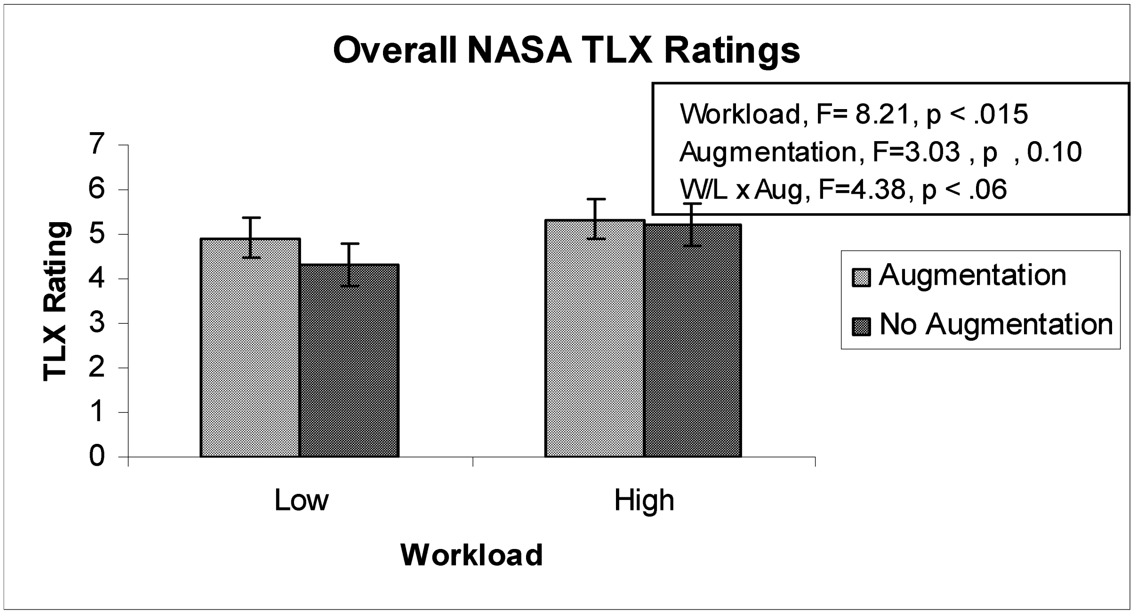


Figure 14. Overall TLX workload rating.

3.5.4 Mitigation Behavior Analysis

3.5.4.1 Overview

The central question with regard to the CWA gauges is the following: How effective was the CWA in driving the mitigation strategies? The Communications Scheduler enhanced performance during the CVE to augment communications management. These performance enhancements are the result of two components:

1. Mitigation strategies that are effective in improving performance, and
2. A CWA that measures cognitive state in a meaningful way in order to effectively drive the mitigation strategies.

The performance-improving characteristics of both needed to be established in isolation in order to interpret any performance improvement. The effectiveness of the mitigation strategies to improve communication management performance was established in a behavioral study (see Whitlow et al., 2004). Building on the knowledge that the mitigation strategies would improve performance if they were driven by the CWA correctly, the AugCog team can now interpret the performance results of the CVE to establish that CWA indeed does drive the mitigation strategies effectively, and thus was producing a meaningful measure of cognitive state.

3.5.4.2 Mitigation Behavior Analysis

The Communications Scheduler leveraged the context, the CSP, and the message characteristics to decide what actions to perform. Messages were characterized by priority (low, medium, or high). The Communications Scheduler looked at three gauges before a message was presented: Engagement, Arousal, and Stress. Each gauge could

have a value of high, medium, low, or unknown. Based on the combination of gauges, the Communications Scheduler performed one of four actions when deciding how to first present the message:

- Present (Audio, Normal)—presented the message immediately in the audio modality with the appropriate "normal" tone preceding it.
- Present (Audio, Escalate)—presented the message immediately in the audio modality with the appropriate "higher saliency" tone preceding it.
- Present (Text)—presented the message immediately in the text modality.
- Not Presented—deferred the message for presentation after the mission is complete (i.e., do not play the message during the mission).

After a message was presented, the Communications Scheduler looked at the XLI and P300 novelty detector gauges to determine if the participant had the attentional resources at the moment of message presentation to properly attend to and understand the message. Based on the combination of the two "after" gauges, the Communications Scheduler performed one of four actions:

- Replay (Audio, Same)—Replayed the message immediately in the audio modality with the same tone used previously preceding it.
- Replay (Audio, Escalate)—Replayed the message immediately in the audio modality with a higher, more salient tone than used previously preceding it. Note that if the first presentation was of the "higher" tone, this replay would use the "highest" tone.
- Done—Did nothing, as the gauges had sensed that the participant comprehended the message.
- Not Applicable—Did nothing, as the "before" decision of the Scheduler precluded any need to make an "after" decision.

Table 4 summaries the actions taken before and after message presentation by the Communications Scheduler for high-, medium-, and low-priority messages. The table was broken out into the actions taken during the low-workload scenarios, the high-workload scenarios, and all scenarios.

Table 4. Actions taken by the Communications Scheduler during the CVE.

	BEFORE First Message Presentation				**AFTER** First Message Presentation			
Low-Workload Scenarios								
Action:	Present (Audio, Normal)	Preset (Audio, Escalate)	Present (Text)	Not Presented	Replay (Audio, Same)	Replay (Audio, Escalate)	Done	Not Applicable
Priority High	112	116	0	0	0	114	114	0
Priority Med	144	0	0	0	81	0	61	0
Priority Low	0	0	98	10	0	0	0	108
TOTAL	256	116	98	10	81	114	175	108

High-Workload Scenarios								
Action:	Present (Audio, Normal)	Preset (Audio, Escalate)	P(Text)	Not Presented	Replay (Audio, Same)	Replay (Audio, Escalate)	Done	Not Applicable
Priority High	83	133	0	0	0	105	111	0
Priority Med	120	0	0	0	67	0	53	0
Priority Low	0	0	107	25	0	0	0	132
TOTAL	203	133	107	25	67	105	164	132
All Workload Scenarios								
Action:	Present (Audio, Normal)	Preset (Audio, Escalate)	P(Text)	Not Presented	Replay (Audio, Same)	Replay (Audio, Escalate)	Done	Not Applicable
Priority High	195	249	0	0	0	219	225	0
Priority Med	264	0	0	0	148	0	114	0
Priority Low	0	0	205	35	0	0	0	240
TOTAL	459	249	205	35	148	219	339	240

The Communications Scheduler rule set considered message priority along with the CWA gauge values when deciding what actions to take. Thus, some of the cells in Table 4 are zero because the Communications Scheduler would never, for instance, escalate the tone for medium- or low-priority messages. Similarly, high-priority messages were always presented.

For the "before" actions, high-priority messages were escalated in tone more often for the high-workload scenarios (61.5%) than for the low-workload scenarios (50.8%). Incoming audio low-priority messages were either presented as text or not presented at all. For the low-workload scenarios, 90.7% of the messages were changed to the text modality, and 9.3% were not presented. In the high-workload scenarios, 81.1% of the messages were presented as text, and 18.9% were not presented. The "after" actions showed a more even spread between the actions taken for low- and high-workload scenarios. Across all scenarios, the actions by the Scheduler are summarized in Table 5.

Table 5. Distribution of "before" and "after" actions for low- and high-workload scenarios.

	BEFORE First Message Presentation				**AFTER** First Message Presentation			
Action:	Present (Audio, Normal)	Preset (Audio, Escalate)	Present (Text)	Not Presented	Replay (Audio, Same)	Replay (Audio, Escalate)	Done	Not Applicable
Low Workload	53.3%	24.2%	20.4%	2.1%	16.9%	23.8%	36.6%	22.6%
High Workload	21.4%	14.0%	11.3%	2.6%	14.3%	22.4%	35.0%	28.2%
All Scenarios	48.4%	26.3%	21.6%	3.7%	15.6%	23.2%	35.8%	25.4%

3.5.4.3 Acknowledgments

Each trial in the CVE contained, on average, ten messages. About five to seven of these messages required an overt response from the participants—either an acknowledgment that they heard *and* understood the message or a response with some relevant information. For these messages, the experimenter recorded whether a participant responded appropriately. In addition, if a message was repeated, the participant had another opportunity to respond. Summaries of the number of times that messages were acknowledged (more precisely, participant responded appropriately) upon first and second presentation are presented in Table 6. The participant acknowledged the first presentation of the message 72.5% of the time, therefore failing to respond 27.5% of the time. In cases where the participant did not respond appropriately, the message was repeated 71.6% of the time (with a follow-up acknowledgment rate of 95%). In cases where the participant did acknowledge the first presentation of the message, the message was repeated anyway 23.7% of the time.

Table 6. Communications Scheduler actions with regard to participant acknowledgment of messages.

Message Presentation	First Presentation	Message Played Once Only	Message Repeated, Participant Then Acknowledges	Message Repeated, Participant Does **Not** Acknowledge
Participant acknowledges	408 (72.5%)	309	97 (23.7%)	2
Participant does **not** acknowledge	155 (27.5%)	38	111 (71.6%)	6
TOTAL	563	347	208	8

3.5.4.4 Gauge Analysis

The Communications Scheduler kept a log file of all the decisions it made and the gauge values it received from CWA at the time it made a decision. Thus, the Communications Scheduler logs recorded the gauge values at the most important times of interest, namely, when a mitigation decision was made. This section summarizes the gauge behavior at those decision times. Note that the analysis of gauge values here is not based on a complete set of data for all times during the trial. Rather it is based on samples of the gauges at semi-random times (i.e., when messages arrived).

The first question asked was: Were the gauges outputting random, evenly distributed, values? There were three "before" gauges, or gauges that were looked at before a message was presented to determine how to present it. They were the Engagement, Arousal, and Stress gauges. Table 7 presents the counts of the three "before" gauges for the low- and high-workload scenarios, as well as the total count for all scenarios. Clearly, the Engagement gauge primarily returned values of medium for all scenarios. The distribution was 12.3% high, 86.3% medium, 0.7% low, and 0.6% unknown. The Arousal gauge was more evenly distributed, while favoring medium and low. The distribution was 18.2% high, 36.2% medium, 34.9% low, and 10.6% unknown. Finally, the Stress Gauge was most often medium or high. The distribution was 22.6% high, 70.8% medium, 6.6% low, and 0.0% unknown.

Table 7. Counts of "before" gauges for low- and high-workload scenarios.

Gauge	Engagement				Arousal				Stress			
Gauge Value	High	Med	Low	Unk	High	Med	Low	Unk	High	Med	Low	Unk
Low workload	162	1088	8	10	236	428	462	142	281	915	72	0
High workload	136	996	10	4	203	447	381	115	265	793	88	0
All scenarios	298	2084	18	14	439	875	843	257	546	1708	160	0

Each of the three "before" gauges had four possible combinations: high, medium, low, and unknown. Thus, there were 64 possible combinations of gauge values. If the gauges were outputting random values, it was expected that any of the 64 combinations would be equally likely. Figure 15 is a histogram of the 64 possible combinations of "before" gauges. Only 34 combinations actually occurred, and the histogram clearly showed that the distribution was not even. The most common combination (523 of a possible 2414, or 21.6%) of gauges was medium-medium-medium, as would be expected with a normalized set of gauges. In fact, the top 15 combinations accounted for 95% of the occurrences. The Engagement gauge was medium for the top seven combinations, or for a total of 79% of the decision points. Arousal varied among all three levels for the most common combinations. Stress varied between medium and high for the top nine combinations, or for 89% of the trials.

The same analysis on the "after" gauges (XLI and P300) reveals a similar nonrandom distribution of gauge combinations. Table 8 presents the counts of the two "after" gauges for the low- and high-workload scenarios, as well as the total count for all scenarios.

The XLI gauge was distributed among all three values. The distribution was 28.6% high, 27% medium, 30.7% low, and 13.6% unknown. Note that there was a considerable number of times when the XLI gauge was "unknown." The XLI functions by surveying a "watch window" that started halfway through a message presentation and continued for 500 milliseconds after a message ends. If during this watch window the XLI updated with a value, then it was compared with a previous value (before message presentation) to establish whether executive load has increased, decreased, or stayed the same. However, if the XLI value fell within the watch window but a second message was already playing, there was no way to disambiguate the XLI reading as to which message it was referring to. This occurred 13.6% of the time. This was a good example of why the Communications Scheduler rule set must be written to account for instances when a gauge value cannot be interpreted as meaningful (and thus "unknown"), even though it has a value.

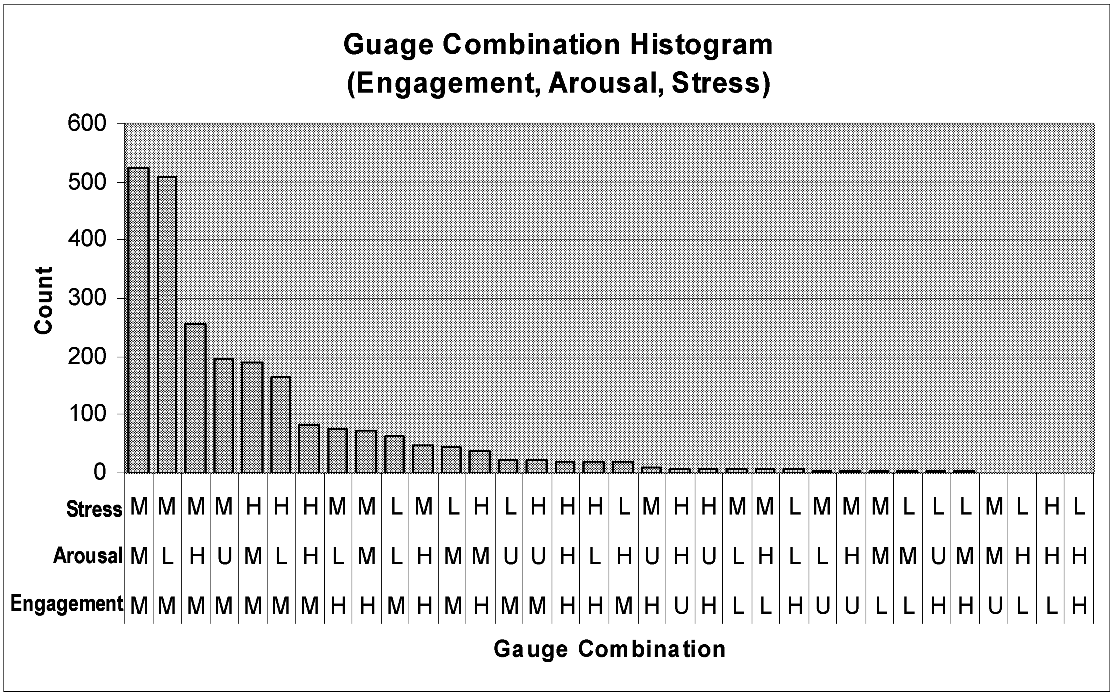


Figure 15. Gauge combinations histogram for the "before" gauges: Engagement, Arousal, Stress.

Table 8. Counts of "after" gauges for low- and high-workload scenarios.

Gauge	XLI				P300 Novelty Detector			
Gauge Value	High	Med	Low	Unk	High	Med	Low	Unk
Low Workload	152	159	165	85	62	39	49	411
High Workload	153	129	162	60	55	37	41	371
All Scenarios	305	288	327	145	117	76	90	782

The P300 gauge registered high more often than medium or low, with the vast majority of the values registering as unknown. The distribution was 11.0% high, 7.1% medium, 8.5% low, and 73.4% unknown. Since P300 only gave a value when prompted for high and medium priorities, most of the time the gauge did not have a value.

The histogram of all the 16 possible "after" gauge combinations is illustrated in Figure 16. Not surprisingly, the top four combinations are with P300 = Unknown. The XLI gauge varied among all three levels.

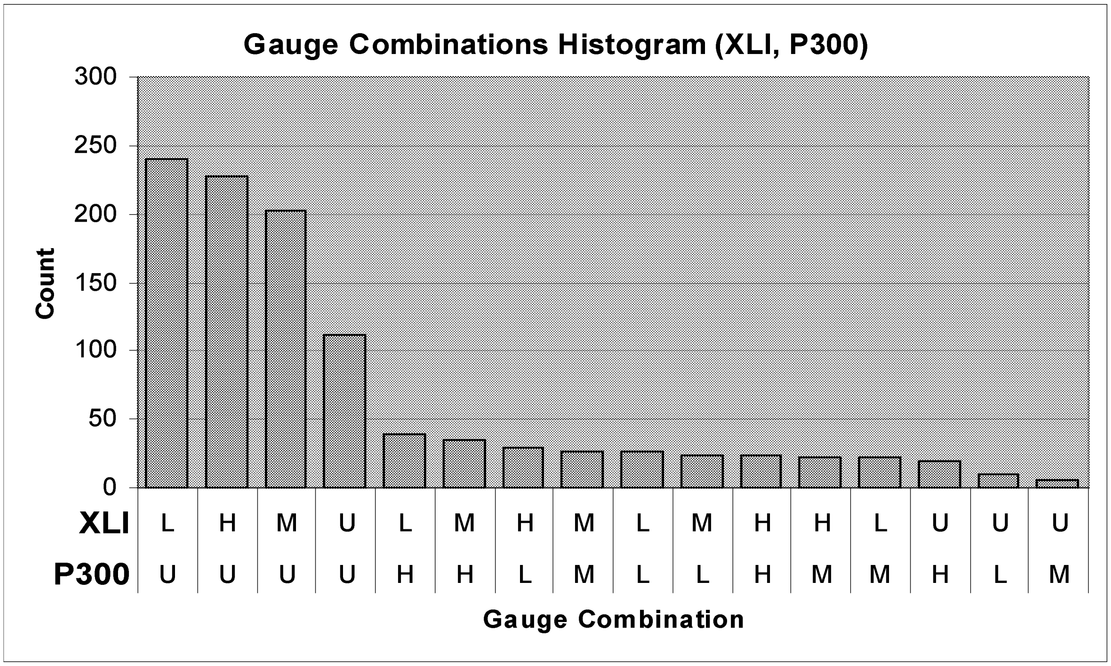


Figure 16. Gauge combinations histogram for the "after" gauges: P300 and XLI.

3.5.5 Qualitative Feedback

Each participant was given a post-experiment questionnaire. The questionnaire was broken into two parts: rating scales and short-answer questions. The results are detailed in Appendix B.

3.6 Phase 2a Discussion

3.6.1 Performance Conclusions

The experimental manipulation produced reliable, significant effects of workload for nearly all performance measures—except for reaction time to OPFOR and message acknowledgment. There were no significant effects of augmentation—except for shooting accuracy and runtime, for which augmentation produced a decrement in performance. In general, it can be concluded that the mitigation strategy did not negatively affect task performance on the primary tasks of IFF or navigating through the environment. As for communications management, the secondary task that was the focus of mitigation, there was a numerical (though not significant) effect for both situation awareness and message acknowledgment under high-workload conditions and both workload conditions, respectively; moreover, the mitigation strategy reduced the high-workload decrement for message acknowledgment.

3.6.2 Mitigation Response

The mitigation strategies were designed to improve performance on the communications management task while not decrementing performance on the Navigate to Objective and IFF tasks. A behavioral study established that the mitigation strategies, if applied correctly, should improve performance as intended. The CVE results showed performance results as described above. In addition, an analysis of the behavior of the

system with regard to participant acknowledgment and comprehension of messages was compelling. In cases where the participant did not respond appropriately to an incoming message requiring a response, the system repeated the message 71.6% of the time (with a follow-up acknowledgment rate of 95%). In cases where the participant did acknowledge the first presentation of the message, the message was repeated anyway 23.7% of the time. Thus, based on cognitive state, the system was able to infer a participant's message comprehension and repeat unattended messages in the majority of cases, with a substantially lower false alarm rate.

3.6.3 Subjective Ratings

Overall, the mitigation strategy did produce higher workload for the participants, as indicated by the marginally significant effect of augmentation ($F = 3.03$, $p < .10$) for overall NASA TLX ratings. This was likely a result of repeating messages for the augmented conditions; however, this result was primarily driven by the difference between Augmentation ON and OFF under low workload. Comparison of the low- and high-workload conditions revealed an interesting pattern suggesting that augmentation either eliminated or reduced this difference for high workload—especially for the Effort and Mental Demand subscales as well as the Overall NASA TLX ratings. This suggested that as workload increases, the mitigation strategy became more valuable in managing perceived workload.

3.6.4 Gauge Correlations

The significant positive correlations between average gauge correlations for all participants indicated that not only did the CLIP have redundant measures that were sensitive to experimental manipulation, but that it detected both neurophysiological and physiological response to task stress. This was also an indication that some gauges might be measuring similar prolonged states that would be represented at the trial-wide level. Furthermore, this was an indication that the task environment taxes both the cognitive and physiological resources of participants. The fact that all of the significant correlations occurred in the high-workload conditions indicates that the current suite of gauges reliably detected overload conditions. This was not surprising, given the nature of the task environment, which challenged the perceptual and cognitive resources of participants. Since operations could also be compromised by hazardous under-load conditions, such as boredom, inattention, drowsiness, or daydreaming, future work should include slower paced, longer duration tasks within the AugCog environment to assess gauge response to underload.

Several gauge combinations both positively and negatively correlated in a more moment-by-moment timeframe. This suggests that the gauge suite also had a more transient responsiveness to immediate task requirements. Also of interest were the occasional, but significant, negative correlations between gauges. The stress and engagement gauges for participant 5 were positively correlated in trial 5, and were negatively correlated in trial 7. This suggested a more complex dynamic between gauges, and further analysis might identify those circumstances where the physiological and neurophysiological gauges diverge.

4 Augmented Cognition Program Phase 2b

4.1 Phase 2b Introduction

4.1.1 Phase 2b Research Team

The Honeywell AugCog team in phase 2b consisted of the collaborative efforts of Honeywell Laboratories, Carnegie Mellon University (CMU), City College of New York (CCNY), Clemson University, Columbia University, Human Bionics, Institute of Human and Machine Cognition (IHMC), Oregon Health and Sciences University, and UFI. This team has developed the Augmented Cognition (AugCog) system for application to the U.S. Army's Future Force Warrior (FFW) program. In addition, the team was advised by the Natick Soldier Research, Development and Engineering Center (NSRDEC). Phase 2b of the program encompassed work done between January 1, 2004, and December 31, 2004.

4.1.2 Phase 2b Research Objectives

Honeywell was charged with addressing the attention bottleneck in joint human-machine system performance. The proposed research aimed to validate the applicability of established noninvasive neurophysiological and physiological state detection techniques in a virtual environment that represented dismounted Soldier combat operations, and showed significant performance improvement of a joint human-automation system employing mitigation strategies triggered by the aforementioned assessment of cognitive state.

The appropriate allocation of attention is important to the Army and the FFW program because it directly affects two cornerstone technology thrusts within the program: netted communications and collaborative situation awareness. The application of a full range of netted communications and collaborative situation awareness (SA) will afford the Future Force Warrior (FFW) unparalleled knowledge and expand the effect of the Future Force three-dimensionally. Task analysis interviews with existing military operations identified factors that negatively affect communications efficacy. In one example, in the first few minutes of any intense mission, radio communications were a suboptimal method of communications because everybody was intensely focused on the tasks at hand. In one famous raid, for example, the commander did not hear the radio communications informing him that the plan had changed until he was physically grabbed by the ground force commander and given this critical information. The commander responded by radioing his own troops, who also did not respond. The implications of these kinds of situations are many, but first and foremost, mission-critical information must be reliably communicated. What aspects of the communication method can be altered to improve the chances that a message was received and understood? Does it require a multimodal, physical alert? Should communications be limited to only critical messages during high-workload situations?

The Honeywell team has developed a set of cognitive gauges based on real-time neurophysiological and physiological measurements of the human operator.

The virtual environment (VE) test-bed facilitated the creation and evaluation of cognitive gauges for determining cognitive workload. Cognitive workload was (broadly) defined as the amount of mental effort needed to perform satisfactorily on a task. Based on neuro-physiological and physiological states, these gauges were used to drive an adaptive cognitive assistance system for dismounted combat operations. With the aid of the proposed adaptive system the team hoped to increase the Soldiers' situation awareness, survivability, performance, and information intake by improving their ability to comprehend and act on available information. It was hypothesized that this adaptation of the Soldier's workspace would lead to greater joint human-automation performance in dismounted Soldier operations. It was anticipated that the Augmentation Manager (AM) would help manage the incoming information by scheduling the communications to be received by the Soldier at the most optimal period, offloading tasks or portions of tasks to automation when the Soldier is overwhelmed, and providing information in multiple modalities (audio, visual, tactile) to ensure comprehension. A high task load condition prompted the automation to defer all but the highest priority messages, offload tasks, or change the modality of information presentation; a low-load condition indicated an appropriate time for interruption and higher levels of Soldier participation in ongoing tasks. Without these mitigations, the Soldier became overloaded with information and had to decide when and where to focus attention among the myriad high-priority communications and high-priority tasks.

4.1.3 Phase 2b Experiment Plan

For Phase 2b, Honeywell conducted two separate Concept Validation Experiments (CVEs). The first CVE was held at the Institute for Human and Machine Cognition (IHMC) and focused on the development of a wide range of mitigation strategies in a militarily realistic virtual environment. The second CVE, held at the CMU Motion Capture (MoCap) laboratory, focused on the ability to detect cognitive state in a (semi-) mobile VE. These environments were chosen because of the flexibility they offered in creating operationally realistic scenarios. These environments also provided the ability to manipulate the attentional demands associated with tasks. Situating tasks within these VEs allowed the AugCog team to precisely relate simulation events to neurophysiological states assessed by the gauges. The two VEs also provided insight into the performance of the gauges under different levels of mobility.

4.2 Phase 2b Attention Bottleneck

An approach was adopted that considered the joint human-computer system when identifying bottlenecks to improve system performance. Key cognitive bottlenecks constrain information flow and the performance of decision-making, especially under stress. From an information-processing perspective, only a limited amount of resources can be applied to processing incoming information due to cognitive bottlenecks (Broadbent, 1958; Treisman, 1964; Kahneman, 1973; Pashler, 1994). The DARPA AugCog program identified four key cognitive challenges related to different components of information processing: 1) the sensory input bottleneck, 2) the attention bottleneck, 3) the working memory bottleneck, and 4) the executive function bottleneck (Raley, Stripling, Schmorrow, Patrey, & Kruse, 2004). The Honeywell team focused primarily on the attention bottleneck, although the other bottlenecks were addressed in the studies described herein. Many varieties of attention were considered to optimize their

distribution (Parasuraman & Davies, 1984): executive attention, divided attention, focused attention (both selective visual attention and selective auditory attention), and sustained attention. Breakdowns in attention lead to multiple problems: failure to notice an event in the environment, failure to distribute attention across a space, failure to switch attention to highest priority information, or failure to monitor events over a sustained period of time. A simplified hierarchy of the component dimensions of attention is shown in Figure 17.

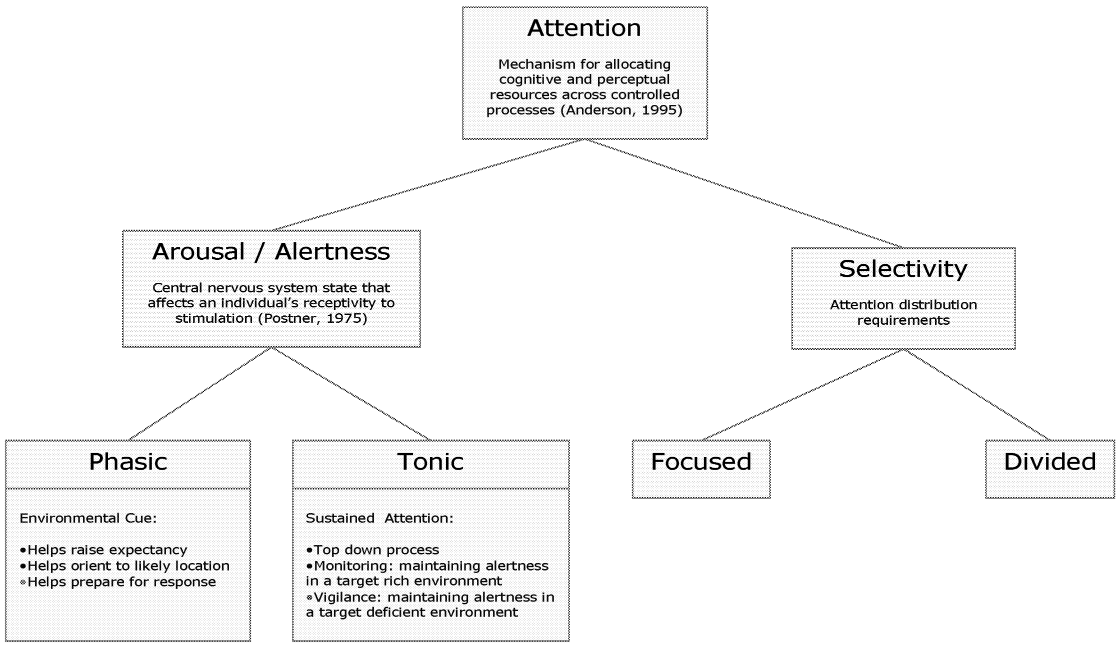


Figure 17. Simplified hierarchy of attention.

Attention can be broadly defined as a mechanism for allocating cognitive and perceptual resources across controlled processes (Anderson, 1995). To perform effectively in military environments, one must have the capacity to direct attention to task-relevant events in a dynamic environment (*alertness*). Additionally, one must be able to narrow or broaden one's field of attention appropriately depending on the demands of a task (*selectivity*). Attention can be stimulated by external events (*phasic attention*), e.g., reacting to gunshots, or a loud aural warning. Alertness can also be maintained consciously, as a controlled top-down process (*tonic attention*). Examples of tonic attention include remaining vigilant while screening baggage at a security checkpoint or looking for insurgents from surveillance positions over a span of hours. While phasic attention is mostly instinctive and automatic, tonic attention requires active effort on the part of a person. A vast body of literature attests to the difficulty of maintaining tonic attention over prolonged periods (e.g., Cabon, Coblentz, Mollard, & Fouillot, 1993; Colquhoun, 1985). Tonic attention was an area of focus in the research reported here. The team explored the use of gauges to detect and drive mitigations during periods when tonic attention levels may be inadequate. This was done in the context of a vigilance task to be described later.

Selectivity is another dimension of attention that is critical for task performance. Warfighters have to be able to distribute their attention over information sources

effectively in order to accomplish various tasks. Attention has to be highly *focused* in many task contexts. Examples include a bomb disposal expert tuning out distractions to carry out intricate procedures associated with deactivating an incendiary device, or a sniper taking aim at a target. However, many tasks require attention to be *divided* across a diverse range of information sources. This is particularly true in today's information centric warfare environment where the Warfighter must attend to potentially hostile events around him or her while maintaining communications and interacting with a range of information devices. An emphasis of the research reported here was performance under conditions where limited attentional resources have to be distributed widely in order to perform effectively. Several of the experiment scenarios to be discussed later explored the efficacy of gauge-driven mitigations under divided attention demands.

4.3 IHMC CVE System Design and Architecture

Details of the Closed-Loop Integrated Prototype (CLIP) configuration used in phase 2b can be found in Appendix C. A brief overview of the principal components is provided in this section.

4.3.1 Cognitive State Classification

Given the hierarchy of the attention bottleneck presented above, the possible suite of cognitive gauges was assessed to determine their appropriateness for the CVE. This section summarizes each gauge and assesses whether the cognitive constructs it measures are meaningful measures of the divided attention paradigm and/or the vigilance paradigm. Given the conclusions described in this section, gauges were chosen for each paradigm and tested within the CVE. The gauges selected were expected to detect extremes under which attentional resources may be inadequate.

4.3.1.1 Engagement Index Gauge

The Engagement Index was a ratio of electroencephalogram (EEG) power bands (beta/(alpha + theta)). The Engagement Index, as described by Freeman et al. (1999) was a measurement of how cognitively engaged a person is in a task, or the level of alertness. Adaptive systems have used this index to drive control of the automation between manual and automatic modes. In fact, the index has been used to successfully control an automation system for tracking performance and a vigilance task (Freeman et al., 1999; Mikulka, Scerbo, & Freeman, 2002; Pope, Bogart, & Bartolome, 1995).

Prinzel et al. (1999) reported that adaptive task allocation may be best reserved for the endpoints of the task engagement continuum. Therefore, two levels of engagement (low, high) were measured in both of the studies described here. The Engagement Index reflected the selection and focus on some aspect at the expense of the other competing demands; thus it was a measure of focused attention. High levels of engagement reflected selection and attentional focus, whereas lower levels of engagement indicated that the participant was not actively engaged with some aspect of the environment.

In the current operational environment, the Engagement gauge tracked the ability to sustain tonic attention over a period of time and was particularly sensitive to low-workload conditions. Thus, it was most appropriate to the sustained attention (vigilance)

paradigm in the IHMC evaluation or the task-shedding period after the status report in the CMU CVE, although it may also be appropriate for the divided attention paradigm.

4.3.1.2 Stress Gauge

IHMC developed a composite *Stress Gauge* (Raj et al., 2003; Kass et al., 2003). The gauge used a weighted average of the three inputs (Video Pupilometry (VOG), High Frequency Electrocardiogram (HFQRS ECG), and Electrodermal Response (EDR) to detect the participant's response to changes in cognitive load within the virtual environment. The gauge was used to detect cognitive stress related to managing multiple competing tasks on a moment-to-moment basis.

The Stress Gauge was only tested in the IHMC study. In this current operational environment, it tracked the autonomic response to time pressure in a high-workload environment; thus, it was appropriate for the divided attention paradigm.

4.3.1.3 Arousal Meter Gauge

Clemson University's Arousal meter (Hoover & Muth, 2003) derived autonomic arousal from the cardiac interbeat interval (IBI), derived from the electrocardiogram (ECG) at 1-millisecond accuracy. The gauge had three levels (low, medium, and high). Increases in this score were associated with increased autonomic arousal and decreases with decreased autonomic arousal.

In the Phase 2b CVE, the Arousal Gauge tracked decrements in performance due to low arousal states and thus was most appropriate to the vigilance paradigm, although it showed some promise for detecting changes in workload in a divided attention paradigm.

4.3.1.4 eXecutive Load Index Gauge

Human Bionics developed a gauge called the *eXecutive Load Index (XLI)* (DuRousseau, 2004, 2004b) to measure patterns in tightly coupled cortical networks tied to an individual's allocation of attentional resources as one's cognitive state changes in response to conditional task load. The index was designed to measure real-time changes in cognitive load related to the processing of messages. This gauge was previously validated to discern trial difficulty in a continuous-performance high-order cognitive task battery.

In the Phase 2b IHMC CVE, the XLI gauge tracked the active inhibition of competing tasks and thus was most appropriate to a divided attention paradigm. At the CMU CVE, the XLI was reformulated to look at workload.

4.3.1.5 P300 Novelty Detector Gauge

The EEG Auditory P300 reflected a central nervous system response to behaviorally relevant infrequent sounds. Previous literature (Wickens, Heffley, Kramer, & Donchin, 1980) suggested that P300 amplitude in response to a task-relevant infrequent auditory stimulus is modulated by attentional resources: If the participant was very focused on a primary task, the auditory stimulus would be missed and the corresponding P300 diminished. Columbia University and CCNY created a gauge called the *P300 novelty detector (*Sjada, Gerson, & Parra, 2003) that spatially integrated signals from sensors

distributed across the scalp, learning a high-dimensional hyperplane for discriminating between task-relevant (incoming message auditory alert) and task-irrelevant responses.

The P300 gauge was integrated into the IHMC environment only. In this operational environment, a tone was played before an auditory message to evoke a P300 activity. Mitigation strategies were based on the assumption that the presence of a strong evoked response indicated that participants have sufficient attentional resources to process the incoming message. The gauge included frontal and parietal electrodes. The P300 gauge tracked the attentional resources to attend to novel stimulus and was also an indirect measure of the response capacity to competing tasks and attentional narrowing. Thus it was most appropriate to the divided attention paradigm.

4.3.2 Mitigation Strategies for IHMC CVE

Four principal mitigation strategies were employed by the AM, each addressed a different task found in the dismounted Soldier domain: Communications Scheduling, Medevac Negotiation, Tactile Navigation Cueing, and Mixed-Initiative Target Identification.

4.3.2.1 Communications Scheduler

Honeywell developed the Communications Scheduler to mitigate divided attention tasks via task-based management and modality-appropriate information presentation strategies. Of particular importance was the Soldier's ability to handle continuous inflow of netted communications and directing his or her attention to the highest priority task to complete his/her mission in this highly dynamic environment. This was crucial not only to the Soldier's own survival but also to that of his/her fellow Soldiers (Dorneich, Whitlow, Ververs, Mathan, et al., 2004b).

The system was tasked with determining when and how information was displayed to the Soldier. The Communications Scheduler scheduled and presented messages to the Soldier based on the cognitive state profile (CSP) (derived from the gauges), the message characteristics (principally priority), and the current context (tasks). Based on these inputs, the Communications Scheduler passed through messages immediately, deferred and scheduled non-relevant or lower-priority messages, escalated higher priority messages that were not attended to, diverted attention to incoming higher priority messages, changed the modality of message presentation, or deleted expired/obsolete messages.

Messages were characterized by priority (low, medium, or high), depending on how critical they were. There were three priorities with the following definitions:

- High Priority: mission-critical and time-critical
- Medium Priority: mission-critical only
- Low Priority: not critical

At times when the augmentation was in effect, messages were scheduled according to certain rules, as described in Table 9. The action taken by the Communications Scheduler before and after the first message presentation was determined by a cross between the CSP and the message priority.

Table 9. Communications Scheduler rule set.

	Before			After		
Priority	High	Med	Low	High	Med	Low
Workload High	P(audio,higher)	P(text,normal)	P(text,normal)			
Workload Low	P(audio,normal)	P(audio,normal)	P(default,normal)			
Workload Low after High	P(audio,higher)	P(text,normal)	P(text,normal)			
Workload Unknown	P(audio,normal)	P(audio,normal)	P(audio,normal)			
Comprehension High				Done	Done	N/A
Comprehension Low				Replay (up)	Replay (same)	N/A
Comprehension Unknown				Done	Done	N/A

The cognitive state assessor (CSA) determined two CSP decision variables: *Workload* and *Comprehension*. The Communications Scheduler determined the initial message presentation based on a user's current *Workload*. The Communications Scheduler performed one of three actions when deciding how to first present the message:

- Presented the message immediately in the audio modality with the appropriate "normal" tone preceding it.
- Presented the message immediately in the audio modality preceded by the appropriate "higher saliency" tone.
- Presented the message immediately in the text modality on the participant's Tablet PC.

After the first presentation of a message to the user (in audio modality), the Communications Scheduler determined whether to take further action on a message depending on the CSA's assessment of *Comprehension*. Comprehension was an assessment of whether the participant had the attentional resources at the moment of message presentation to properly attend to and understand the message. Based on comprehension, it performed one of four actions:

- Replayed the message immediately in the audio modality preceded by the same tone used previously.
- Replayed the message immediately in the audio modality preceded by a higher, more salient tone than used previously. Note that if the first presentation was of the "higher" tone, this replay would use the "highest" tone.
- Did nothing, as the gauges had sensed that the participant comprehended the message.
- Not Applicable—the "before" decision precluded any need to make an "after" decision.

High-priority messages were mission-critical and time-critical, which means they must have been heard and understood as soon as they arrived. Thus, the Communications Scheduler took the following actions on high-priority messages:

- High-priority messages were preceded by a tone (normal or escalated).
- A visual icon reminded the participant to pay attention (see Figure 18).
- High-priority messages that required an overt response were accompanied by a visual summary.
- Message may have been repeated.

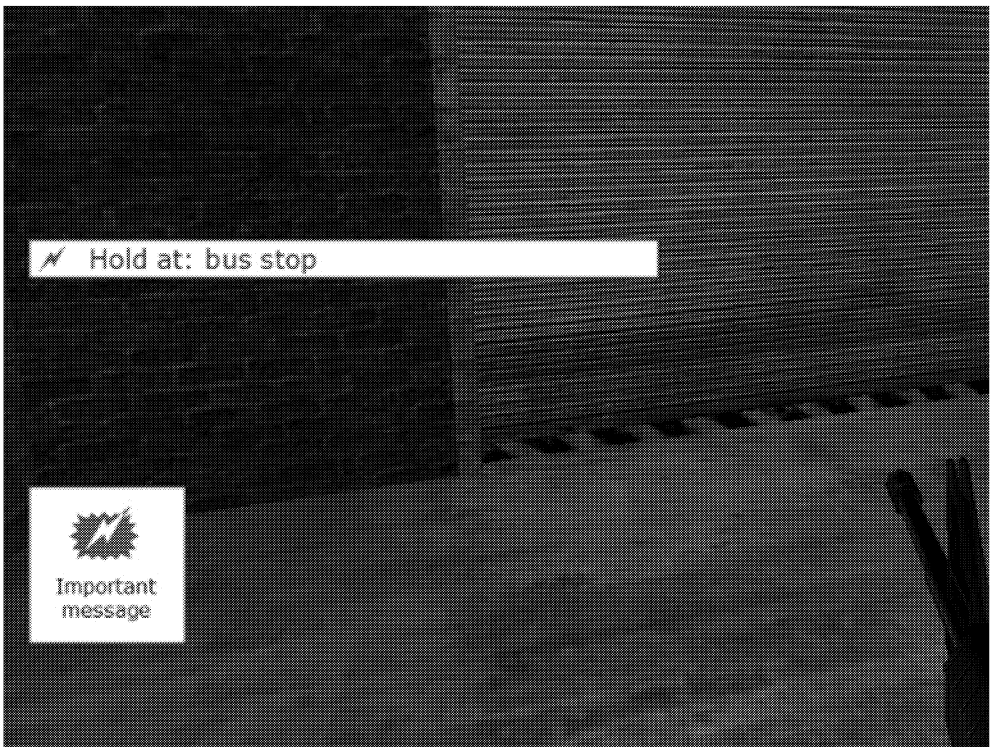


Figure 18. High-priority messages alerted by an icon and (possibly) a text summary on the HUD.

Medium-priority messages were mission-critical but had a larger time window to work with. A medium-priority message was deferred if the system found that the participant was highly engaged in another task. All medium-priority messages were played before the end of the mission. Low-priority messages were not mission-critical or time-critical. They were presented if the participant was not engaged in another task. If the system found that the participant was engaged in another task, the low-priority messages were presented in text format in the message window. Specifically, low- and medium-priority messages were deferred to the Tablet PC application, and a visual icon appeared on the heads-up display (HUD) to alert to the action the scheduler had taken (see Figure 19).

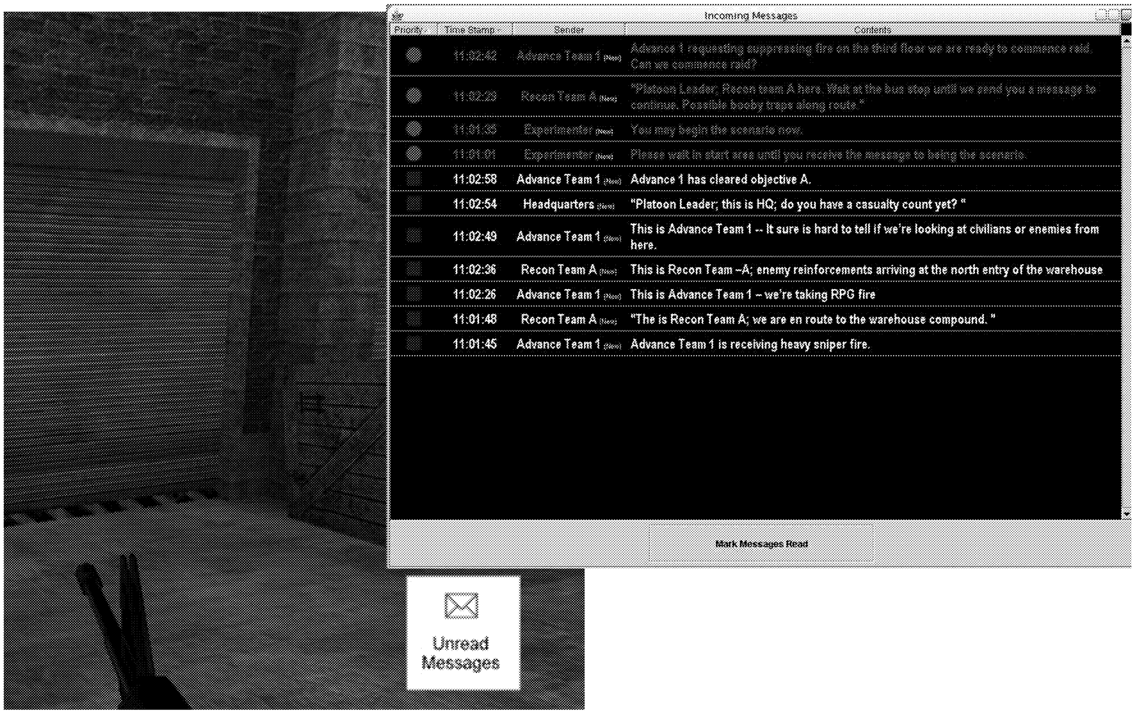


Figure 19. Deferred messages on the Tablet PC (left) with an icon on the HUD (right).

Poorly designed automation can be dangerous. Research shows that unless users are able to predict clearly how an automated system is likely perform, automation may introduce more problems than it solves (Sarter, Woods, & Billings, 1997). The mitigation strategies described here had very clear rules to eliminate uncertainty and unpredictability.

The Communications Scheduler mitigation was invoked when workload was high—for instance, low-priority messages were deferred to the Tablet PC. However, when workload dipped below the threshold used to trigger the message deferral, the Communications Scheduler continued to defer messages. The reason for this was that deferring communications on the basis of moment-to-moment fluctuations in gauge values could have been confusing. Messages could have been misinterpreted without surrounding context if they were played in audio modality after their predecessor messages had been deferred to the Tablet PC (and remained unread for a period of time). If expected messages were not heard, it may have been hard to disambiguate whether this is due to the Communications Scheduler or some mission-related cause. To avoid confusion, once communications scheduling was activated, all low- and medium-priority messages were deferred to the Tablet PC until the user caught up on all messages and clicked a "messages read" button.

4.3.2.2 Tactile Navigation Cueing System

In the unmitigated version of Scenario 2, the participant referred to his or her map on the Tablet PC, oriented him- or herself to the current location, and determined the next-best route to take in order to reach the safe zone while not being ambushed. In the mitigated scenario, the participant received tactile cues that guided him or her in the correct direction to take to reach the safe zone. Thus, the navigation task went from being cognitively intense to one that was essentially reactionary to external stimuli. This was

designed to decrease the task load and cognitive demands, allowing participants to improve performance on the navigation task while not adversely affecting other tasks being done simultaneously. Tactile cues have been shown to be effective in improving performance of spatial tasks, even in the presence of competing secondary workload tasks (Raj, Kass, & Perry, 2000).

The Tactile Situation Awareness System (TSAS) was integrated into the IHMC CVE to provide navigation cueing during mitigated trials. This implementation of TSAS consisted of a 24-tactor belt (2 rows of 12 columns) worn about the upper abdomen of each participant and controlled by the Joint Strike Fighter Tactile Situation Awareness System Laboratory Development Rack. The individual C-2 linear actuator tactors (Engineering Acoustics Inc., Winter Park, FL) were adjusted in pairs to represent the 12 cardinal positions of the clock (12 o'clock centered on the umbilicus). Tactors were fired (using a 300-Hz bipolar sine wave) in pairs to direct participants toward the bearing of their next waypoint or endpoint for cardinal positions and in quads (at one-half amplitude) for "half-hour" positions (e.g., toward 1:30), providing 15 degrees of azimuth resolution. The FFW VE agent continuously returned azimuth, range, elevation, amplitude, and irritability to the TSAS agent. The rate of firing the tactors increased from 1 to 2 to 8 Hz as the participant approached each waypoint. When a waypoint was reached, the VE automatically sent navigation cues relative to the next waypoint until the participant reached the scenario endpoint.

Operationally, pulses from the tactor belt "tugged" the participants in the direction they were expected to go. A redundant visual navigation cue was given via a navigation "bug" (red triangle indicator) on a compass at the top left corner of the HUD. The system was invoked when the CSP indicated *Workload* was high and the participant needed to navigate through an unfamiliar route. However, turning the system off as soon as *Workload* fell below some threshold would leave users disoriented in an unfamiliar area. Thus, once the system is turned on, the navigation mitigation persisted until users reached the safe destination.

4.3.2.3 Medevac Negotiation Agent

The evacuation of injured personnel is a crucial Warfighter function. The task is lengthy and requires a substantial amount of information to be communicated accurately. Performance on this task may suffer under high-workload conditions. Personnel may omit important information or make errors in the information transmitted. Additionally, attention devoted to the medevac information exchange may detract from the performance of other critical tasks. In Scenario 2, the participants navigated through unfamiliar territory while simultaneously coordinating a medevac, both under a severe time deadline. The medevac agent provided the means to offload medevac tasks under high-workload conditions.

The Medevac Negotiation Agent was triggered on the basis of task context and CSP. If *Workload* was high and a medical evacuation (medevac) had to be coordinated, the Medevac Negotiation Application was triggered. A medevac icon on the HUD notified the user about the need to coordinate an evacuation using the medevac agent. The platoon leader (PL) reviewed the medevac information on the Tablet PC and transmitted information using the interactive form. Figure 20 illustrates a Medevac Negotiation

Application presented to the participant on the Tablet PC in the mitigated versions of the scenario. Information available on the FFW Netted Communications network was automatically filled in, and the system presented this information about casualties to the PL for inspection. The system also provided the option of delegating subsequent medevac negotiation to team members facing lower workload demands. Medevac information transmitted using the form was used to organize the evacuation. Any clarification or further negotiation was delegated.

Figure 20. Medevac icon on HUD (right) and Negotiation Application (right).

The Medevac Negotiation Agent only contained the most critical information needed for a medevac. A more detailed information exchange might allow for safer and more efficient medevac operations. Additionally, engaging in medevac transactions may contribute to better situational awareness of a team's status. For these reasons, the Medevac Negotiation Agent was only invoked when the participant's workload was so high and the participant's performance so inadequate that the costs associated with automated negotiation was acceptable in terms of overall performance.

4.3.2.4 Mixed-Initiative Target Identification

Military personnel sometimes have to maintain high levels of sustained attention in environments where target stimuli may be infrequent and hard to detect. An example might be monitoring a camera feed of a compound for the presence of insurgents. Research suggests that performance on these tasks deteriorates considerably over time. Automated systems trained to detect target stimuli in a field may not perform as well as an alert human. Consequently, they may not be able to completely replace the human operator in operational contexts. However, these systems could play a helpful role if they could be triggered when gauges detected a vigilance decrement. This system was modeled after enabling technology (Schneiderman & Kanade, 2004) currently under development. The equipment necessary for such a system would include a display integrated helmet with multispectral vision capabilities. Such a system of mixed-initiative

search with intelligent assistance is part of the FFW vision. For more information, see U.S. Army (2003).

Participants were looking for targets in a series of surveillance photos. The Mixed-Initiative Target Identification System highlighted suspected targets on the surveillance photos, as shown in Figure 21. The system was designed to be an assistant to the human, providing suggestions as to where an enemy Soldier may be hiding. The system detected the presence of an enemy Soldier in a picture and tagged the detected Soldier with a yellow box. However, due to an unacceptable frequency of errors within the system, the participant used the system output for advice but continued to scan. To eliminate ambiguity about whether or not the automated system was providing help with a particular image, assistance was provided in blocks that lasted several minutes.

Figure 21. Mixed-initiative system when automation identifies possible targets.

With a mixed-initiative search, the human performed target detection tasks when the human operator was likely to perform better, and an automated system provided assistance when its performance was likely to be better than a human with a vigilance decrement. The vigilance task consisted of images of a compound being sent to a user at the rate of one every 2 seconds. The images alternated between two perspectives. Participants were asked to signal the presence of enemy Soldiers in the scene. Pilot studies showed alert users were able to detect targets with an accuracy of about 80%. In contrast, following periods ranging from 20 to 40 minutes of sparse targets, performance fell to approximately 40% accuracy. The mixed-initiative target identification process consisted of assistance from a system with a 68% accuracy rate helping out when gauges indicate low attention states. The system's assistance consisted of boxes drawn around areas of the images likely to contain enemy Soldiers. Some common success and failure modes of automated assistance are shown in Figure 22.

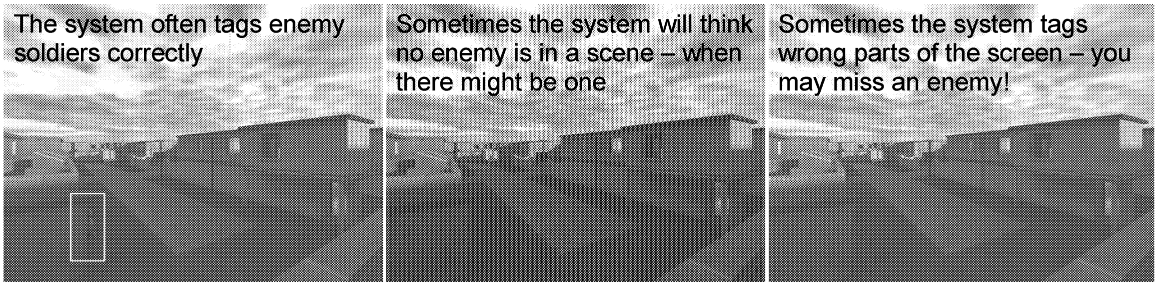


Figure 22. Potential success and failure modes of automated target identification system

The automation was not foolproof. Its accuracy rate of 68% was well below the ideal performance of an alert human. Yet this level of performance is well above chance. While such a system could never replace a vigilant human, it could aid a human who may not be appropriately alert. In addition, other issues such as over-reliance on automation and the human's generally poor ability to passively supervise automated processes precluded automating the process. However, joint human-automation performance during the decremented portion of Scenario 3 was expected to be significantly improved by the mitigation, although not to the level of an ideal human performance. Thus, although one would never employ the automation continuously due to its poor overall performance compared with ideal human performance, there might be times when the human's performance has degraded to the point where assistance from even less-than-perfect automation would significantly improve performance over that of the human alone.

4.4 Phase 2b IHMC Concept Validation Experiment

4.4.1 Experiment Objectives

Several steps were needed to conduct the research necessary to hold the IHMC CVE. Honeywell's approach to the CVE can be framed around a series of research questions that needed to be answered:

1. What attention states should be studied?.
2. What task/scenario features induced those cognitive states, and can scenarios be devised to make them happen? Honeywell performed extensive testing to determine that the scenarios developed put participants at the extremes of workload.
3. Could the gauges correctly identify cognitive states of interest? Each gauge developer validated his/her gauge against data generated in the Pre-CVE.
4. Could the gauges correctly drive the mitigations? Gauge validations were conducted during the Pre-CVE and on the CVE data.
5. Would the mitigations enhance performance? Honeywell conducted behavioral studies with more than 20 participants to ensure that the mitigations, if properly driven by the gauges, would significantly improve performance (see Dorneich et al., 2004b).
6. What level of performance improvement did the mitigations produce? This question was related to what metrics were devised to assess performance.
7. What were the resulting hardware and software requirements to realize the AugCog system?

Figure 23 illustrates these questions in terms of the interactions between the human and the cognitive tasks and mechanisms.

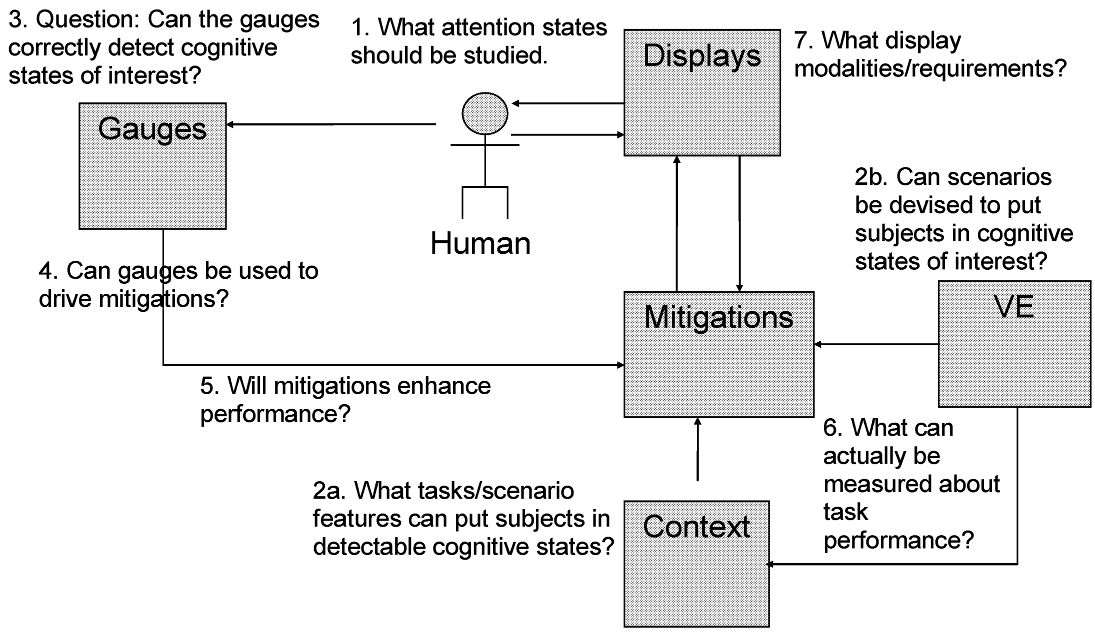


Figure 23. Interactions between the human and the cognitive tasks and mechanisms

4.4.1.1 Expected Results

The approach above listed the series of questions that needed to be answered to meaningfully interpret the data from the CVE. The central question of the CVE was the following: Can the gauges detect the cognitive states of interest consistently enough to drive the mitigations? If the mitigations were driven correctly, performance was expected to improve in the communications tasks in Scenario 1, the navigation and medevac tasks in Scenario 2, and the target detection task in Scenario 3. The goal for the Phase 2b CVE was to attain at least 100% performance improvement on mitigated tasks, with no performance decrement to concurrent tasks and with no negative effect on overall workload.

4.4.2 Operational Scenario

The IHMC CVE focused on two types of attention: tonic arousal required for vigilance and divided attention across multiple tasks. The experiment centered on three tasks (communication, hostile engagement, and navigation) based on input from participant subject experts situated in the FFW program. Maintaining an appropriate level of arousal and engagement during vigilance tasks (tonic attention) such as scouting a reconnaissance duty under stressful and fatigued conditions has always been an issue with the military. In addition, in the information-rich environment provided by the FFW program, the appropriate allocation of (divided) attention is a key to managing multiple tasks, focusing on the most important ones and maintaining situation awareness.

The operational environment for the IHMC CVE was realized in a desktop VE that simulates Mobile Operations in an Urban Environment (MOUT). The VE in all scenarios consisted of a city composed of narrow streets surrounded by two- and three-story

buildings. The environment had an industrial appearance. The visual complexity of the environment contributed to the participant's workload. The participant was faced with a specific number of enemy forces. These forces were presented both at street level and above as snipers. The enemy forces had logic for detecting the presence of the participant or other friendly forces and attacked with varying levels of success (depending on the workload and difficulty settings).

The participant performed all tasks in the environment using a combination of keyboard and mouse controls. The controls allowed the participants to look around the virtual world, to move (walking forward or backward, sidestepping left or right, jumping, and crouching), to shoot their weapons (an approximation of an M16), and to manage messages.

Participants navigated to an objective through familiar and unfamiliar areas. They engaged foes as they navigated to the objective. In addition, the participants managed communication flow between team members and commanders, and supported procedures such as calling for a medevac. Participants sent and received reports, issued and received commands, provided and requested status updates, provided and requested information, and coordinated with friendly forces.

The task environment and scenarios were designed to manipulate the constructs of attention described in the simplified breakdown of attention in Figure 17. Moreover, the scenarios were designed to place individuals in the extremes of the attentional states under study. There were three primary scenarios:

1. Divided attention between communications, engaging foes, and navigation; focused attention on high-priority messages.
2. Divided attention between communications, engaging foes, and navigation; focused attention on high-priority messages.
3. Sustained attention (vigilance) in a target-deficient environment.

4.4.2.1 Scenario 1: Divided Attention

4.4.2.1.1 Description

Scenario 1 focused on three critical task elements of the Raid on Objective mission: Navigate to Objective, Identify Friend or Foe (IFF), and Manage Communications. The participant was a PL, whose goals were to lead the platoon through a hostile urban environment to the objective, while being careful to shoot only enemy Soldiers. Participants also received incoming communications throughout the scenarios. Some messages required an overt, behavioral response. Participants received status updates, mission updates, requests for information, and reports. These incoming communications were a primary source of their SA.

Unlike the Phase 2a CVE, this scenario had a straightforward, simple route. However, the radio communications volume was extremely high. The scenario only included two or three high-priority messages, which told the Soldier to hold at certain locations for a specified amount of time or that the objective location had changed. Failure to heed these high-priority messages caused the participant to encounter an ambush. Figure 24 details the route, the points in the scenario where the high-priority messages occurred, and the potential ambush locations, if the participant failed to heed the messages.

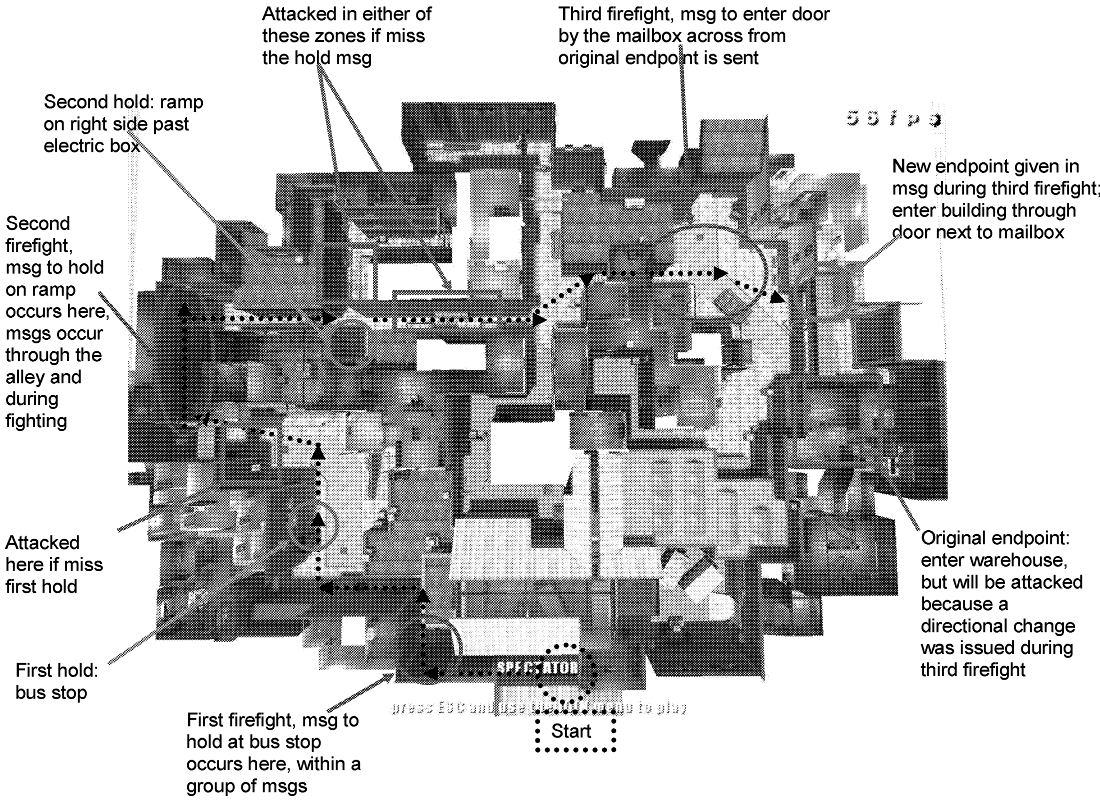


Figure 24. Scenario 1: Divided attention.

The key to Scenario 1 was to put participants into the extremes of workload. When they held at a location for up to 3 minutes, a low task load situation, it was expected that the gauges would register low workload. When participants engaged in a firefight, the gauges were expected to register cognitive states associated with high workload. Validation of the ability of the gauges to distinguish between these periods of low and high workload was an important component in the design of this scenario. Scenario 1 was principally designed to test the performance improvements derived from the Communications Scheduler. Thus, the high-workload times included a high volume of communications traffic to the participants, just at the time when their workload was high due to being targeted by foes. The mitigations utilized this characterization of workload to determine what actions to perform.

4.4.2.1.2 Tasks

Participants' primary responsibility was to ensure their survival while navigating to their objective. The participants had to fight foes. In addition, the participants monitored their radio communications to maintain situation awareness and follow mission commands. Thus, the tasks of this scenario were 1) navigate to objective, 2) engage foes, and 3) manage communications.

4.4.2.1.3 Metrics

For Scenario 1, the metrics of interest were the following.

Message comprehension: For each scenario, there were three messages that required the participants to change their overt behavior. For instance, a participant heard a message to "hold at the bus stop and await further orders." If the participant heard and comprehended the message, he or she would have held at the bus stop. If the participant failed to comprehend the message, he or she would have continued past the bus stop. In this way, the scenario was designed to give clear, unambiguous data on whether a participant comprehended a message. In addition, there were messages that required a participant to respond with a specific piece of information. Note that unlike the CVE in Phase 2a, these messages required more of a response than simply saying "Acknowledge"; rather they required a response with a specific piece of information. A correct response to the query in the message was an indication of comprehension. Thus, the metric was the number of messages a participant correctly responded to (either verbally or through expected behavior).

Situation Awareness: Participants were asked four to eight questions after each scenario to ascertain whether they could recall mission-critical information relayed through the communications. Ability to recall this information was taken as an indirect measure of their ability to build a situation awareness of the mission context. Thus, the metric was the number of situation awareness probe questions correctly answered.

Run Time: Participants were trained on a route to follow. The time it took them to complete such a route while trying not to take hits was a measure of their effectiveness in navigating to the objective.

Hits Taken: The number of hits by opposing force (OPFOR) on the participant was taken as a measure of his or her ability to engage foes.

Hits on OPFOR: The participant's ability to hit OPFOR was taken as a measure of his or her ability to engage foes.

Shooting Accuracy: The percentage of shots fired that hit OPFOR was taken as a measure of the participants' ability to engage foes.

Workload: Participants rated their subjective assessment of workload via the NASA TLX (Task Load Index) scale (Hart & Staveland, 1988). They rated their workload on six rating scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. In addition, these six scales were averaged to produce an overall workload rating.

Qualitative Preferences: Participants were surveyed at the end of the experiment to ascertain the modes (mitigated, normal, or no difference) in which they felt certain tasks were easier.

4.4.2.2 Scenario 2: Divided Attention

4.4.2.2.1 Description

In this scenario, the participant traversed the same initial route as in Scenario 1. However, upon reaching the objective area, the participant was informed that the enemy had set a trap. He or she needed to abandon the objective, get back to the safe zone, and avoid the route he or she just took to the objective. To return to the safe zone, the participant had to

navigate though unfamiliar parts of the city in order to avoid ambushes. The task load stemmed from having to mentally convert an exocentric 2-D representation of an unfamiliar area into an egocentric representation and reason with this newly formed representation. The participants were provided with an updated map on their Tablet PC showing potential ambush zones. Simultaneously, the participants received a request to coordinate a medevac immediately. This was a quite lengthy and communications- and information-intensive procedure. The map and the information requirements of the medevac procedure are illustrated in Figure 25.

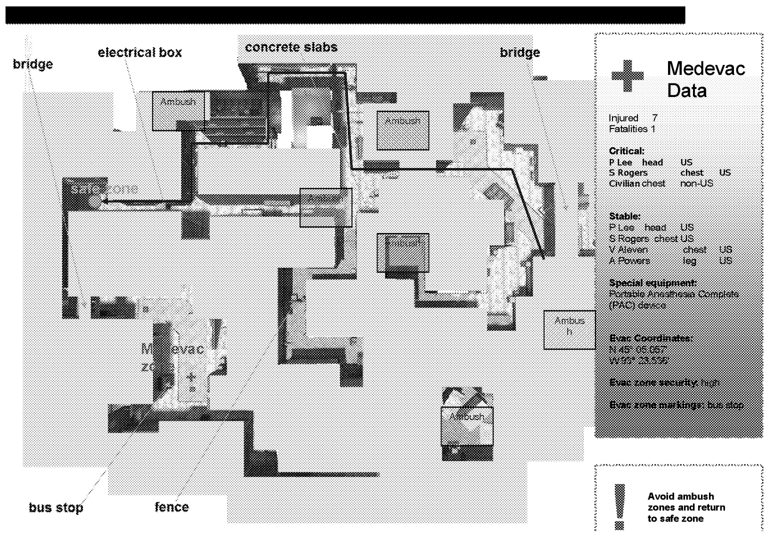


Figure 25. Scenario 2 (divided attention): Tablet PC map and medevac display.

The nominal medevac coordination procedure (see **Table 10**) was simplified to seven questions that were answered by the participant in order to complete the medevac successfully. Both the medevac communications procedure and the navigation to the safe zone task had to be accomplished simultaneously and under extreme time pressure. Additionally, the participant had to engage any foes that he or she encountered.

Table 10. Nominal medevac procedure and modified medevac communications.

Nominal Performance Steps	Communications to Participant
1. Collect all applicable information needed for medevac request.	
a. **Determine the grid coordinates for the pickup site.**	Platoon Leader. We need you to transmit the evacuation coordinates now.
b. Obtain radio frequency, call sign, and suffix.	
c. **Obtain the number of patients and precedence**.	Platoon Leader, Med Team 3. How many need to be evacuated?
d. **Determine the type of special equipment required**.	Platoon Leader, Med again. Do you need any special equipment?
e. **Determine the number and type (litter or ambulatory) of patients**.	Platoon Leader, Med Team 1 again. Can you send us the severity of injuries?
f. **Determine the security of the pickup site**.	Platoon Leader, Med Team 3. Is the evacuation site secure?
g. **Determine how the pickup site will be marked**.	Platoon Leader, can you or your team set off smoke at the evac site when you arrive?
h. **Determine patient nationality and status**.	Platoon Leader, Med Team 3. Are the wounded all U.S. Soldiers?
i. Obtain pickup site NBC contamination information normally obtained from the senior person or medic.	
NOTE: NBC line 9 information is only included when contamination exists.	
2. Record the gathered medevac information using the authorized brevity codes.	
3. Transmit the medevac request.	
a. Contact the unit that controls the evacuation assets.	
(1) Make proper contact with the intended receiver.	
(2) Use effective call sign and frequency assignments from the SOI.	
(3) Give the following in the clear: "I HAVE A MEDEVAC REQUEST"; wait 1-3 seconds for response. If no response, repeat the statement.	
b. Transmit the medevac information in the proper sequence.	
(1) State all line item numbers in clear text. The call sign and suffix (if needed) in line 2 may be transmitted in the clear.	

4.4.2.2.2 Tasks

The mental demands associated with this task were substantial as participants had to split attention among three critical tasks. The participants' primary responsibility was to ensure their survival while navigating in an unfamiliar part of the city to the safe zone, under a time deadline. The participant had to fight foes. In addition, the participant coordinated a lengthy medevac procedure in the same time frame as navigating to the safe zone. Thus the tasks of this scenario were 1) navigation through an unfamiliar area, 2) engage foes, and 3) coordinate medevac.

4.4.2.2.3 Metrics

For Scenario 2, the metrics of interest were the following:

Time to Safe Zone: The time it took a participant to navigate from the warehouse to the safe zone was taken as a measure of the participant's ability to navigate through unfamiliar territory.

Ambushes Encountered: If participants were successful in translating the information on their map display, they should have been able to avoid ambush areas. Thus, the number of ambushes encountered was taken as a measure of the participants' ability to navigate through unfamiliar territory.

Medevac Questions Answered: The number of medevac-related questions the participant was able to answer correctly was taken as a measure of the participant's ability to coordinate a medevac.

Time to Complete Medevac Coordination: The time it took the participant to complete the medevac negotiation process was taken as a measure of the participant's ability to coordinate a medevac.

Hits Taken: The number of hits by OPFOR on the participant was taken as a measure of the participant's ability to engage foes.

Hits on OPFOR: The participant's ability to hit OPFOR was taken as a measure of his or her ability to engage foes.

Workload: Participants rated their subjective assessment of workload via the NASA TLX scale. They rated their workload on six rating scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. In addition, these six scales were averaged to produce an overall workload rating.

Qualitative Preferences: Participants were surveyed at the end of the experiment to ascertain the modes (mitigated, normal, or no difference) where they felt certain tasks were easier.

4.4.2.3 Scenario 3: Sustained Attention (Vigilance)

4.4.2.3.1 Description

In Scenario 3, the participant was a Soldier sitting in the bushes outside a compound with a static view of the compound. The participant was the leader of a reconnaissance unit and was responsible for identifying any targets (enemy Soldiers). The participant received, via his or her Tablet PC, reconnaissance photos from external surveillance cameras. The photos, from two sources, were updated once every 2 seconds. Figure 26 is a surveillance photo shown to the participant on the Tablet PC. Note the presence of a target in the lower left corner.

The experiment protocol for this scenario was a classic vigilance paradigm. The scenario lasted approximately 30 minutes. The first five-minute session had targets occurring at a rate of 14% and served as the measure of baseline performance. This was followed by a 20-minute session with a very low target occurrence rate (3%). This period was expected to produce a vigilance decrement in the participant. The final five-minute session had a target occurrence rate identical to the first five-minute session. Performance in this final session was expected to be decremented.

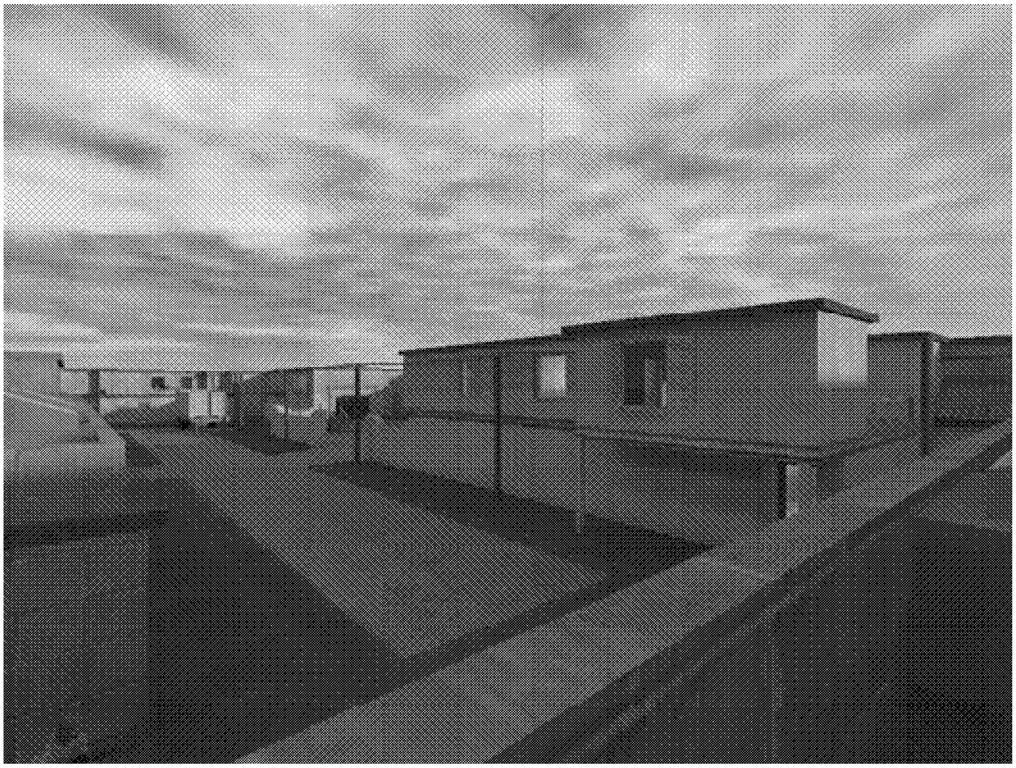


Figure 26. Scenario 3: Vigilance surveillance photo.

4.4.2.3.2 Tasks

In Scenario 3, the participant was asked to perform one task: monitor the surveillance photos to identify targets. Thus, the task was identify targets.

4.4.2.3.3 Metrics

For Scenario 3, the metric of interest was:

Target Identification: The accuracy with which participants identified targets was taken as a measure of performance. Specifically, the accuracy of stage 1 was taken as the (ideal) human baseline performance. The accuracy of stage 3 was the measure of interest as compared to the performance of stage 1. The vigilance paradigm was attempting to induce a decrement in performance between stage 1 and stage 3. Stage 3 then was or was not mitigated, depending on the experimental condition employed for that participant.

4.4.2.4 Relationship Between Mitigations and Scenarios

There are four broad categories of possible mitigations in an AugCog system:

- Task/information management
- Modality management
- Task offloading
- Task sharing

The multiple scenarios of the IHMC CVE provided Honeywell with the opportunity to explore a wide range of possible mitigation strategies.

There were four principal mitigation strategies employed by the AM, each addressing a different task found in the dismounted Soldier domain: Communications Scheduling, Medevac Negotiation, Tactile Navigation Cueing, and Mixed-Initiative Target Identification. Table 11 shows how each mitigation strategy related to the scenarios found in the IHMC CVE. Thus, the scenarios encompassed some form of each of the four broad mitigation categories.

Table 11. Classes of mitigation strategies addressed in the IHMC CVE.

Mitigation Strategy	**Scenario 1** Divided Attention	**Scenario 2** Divided Attention	**Scenario 3** Sustained Attention
Task Scheduling	Communications Scheduler		
Task Offloading		Medevac Negotiation	
Task Sharing			Mixed-Initiative Target Identification
Modality Management	Communications Scheduler	Tactile Navigation Cueing	

In Scenario 1, the primary mitigation was task/information management via the Communications Scheduler. In addition, the Communications Scheduler's ability to change audio messages to text was a form of modality management as well.

In Scenario 2, the system utilized the Medevac Negotiation Tool, a task offloading mitigation, to reduce the workload involved in coordinating the medevac procedure. Medevac coordination was a highly procedural task and thus amenable to offloading. In addition, a tactile display was used via the Tactile Navigation Cueing System, to assist the participant in the navigation to the safe zone task.

In Scenario 3, the system utilized a Mixed-Initiative Target Identification System, as a task-sharing mitigation to improve the performance of the degraded participant. Recall that performance was expected to be severely degraded (as compared with the baseline) in the final five-minute period of the scenario.

4.4.2.5 Mitigations Cost/Benefit Discussion

Although the mitigations described here had the potential for boosting performance when human cognitive resources were limited, they could have had detrimental effects if left on at all times. The benefits and costs associated with these mitigations are shown in Table 12. Gauge-driven mitigation allowed these mitigations to be activated when the benefits were likely to outweigh the costs.

Table 12. Costs and benefits of mitigations.

Mitigation Agent	Benefits	Cost
Communications Scheduler	Allows users to defer responses to messages under conditions when attention has to be split between competing tasks	Loss of momentary situational awareness Lags in responses could break coordination among teams and introduce inefficiencies in the mission
Tactile Navigation Cueing System	With automated navigation assistance, enables users to focus on other critical tasks that demand attention	Loss of situational awareness since user is passive in the navigation task. Cause of many accidents—such as the American Airlines crash in Cali, Columbia.
Medevac Agent	Reduces a lengthy communications exchange to a mouse click	A verbally negotiated medevac reduces ambiguities and possible inefficiencies. It also results in a deeper level of processing of the information which would more likely be recalled at a later time.
Mixed-Initiative Search	Provides assistance in locating targets in visual search tasks	Alert human will perform better on the search task. Leaving the system on all the time could potentially cause users to depend on a suboptimal system.

4.4.3 Experiment Hypothesis

In general, the hypothesis for this experiment was as follows:

- The mitigations will improve performance on the tasks they are mitigating without decrementing other concurrent tasks**.**

Specific hypotheses vary by scenario in the experiment design.

4.4.3.1 Scenario 1: Divided Attention

The CVE hypothesis stated that the "smart" Communications Scheduler would enhance overall performance on the communications management task (as measured by message response and situation awareness metrics) while not significantly degrading performance on the Navigation to Objective and IFF tasks. Specifically, under augmentation, it was hypothesized that participants would have better situation awareness for message content and participants would attend better to high-priority messages.

- *Hypothesis: Better awareness of message content in mitigated condition.*

4.4.3.2 Scenario 2: Divided Attention

The CVE hypothesis stated that performance on the medevac coordination would improve due to the accurate communication of critical medevac information. Performance in the navigation task could be improved because cognitive resources that used to be shared with the communications tasks were offloaded and thus available. Navigation performance was aided by tactile perceptual cues. Tugs from a tactor belt guided users to a safe zone while avoiding ambushes.

- *Hypothesis: Accurate transmission of medevac information with gauge-driven mitigation. Safer (i.e., fewer ambushes) and efficient (i.e., faster) navigation with mitigation.*

4.4.3.3 Scenario 3: Vigilance

The CVE hypothesis stated that performance during the degraded portion of the scenario (i.e., final 5 minutes) in the mitigated case would be better than performance of the corresponding portion of the unmitigated scenario. In addition, while joint human-automation performance in the mitigated case would show improvement over the unmitigated performance, it was not expected to exceed performance of the human alone when the human is alert (i.e., the baseline portion of the scenario). Performance was measured in terms of the target detection accuracy in classifying images as containing images or not.

- *Hypothesis: More accurate* target detection *performance in decremented periods with gauge-driven mitigation.*

4.4.4 Experiment Design

There were two independent variables:

- Mitigation (on/off)
- Scenario (three, which vary by attention type)

The study consisted of three two-factor experiments. Each experiment compared performance under gauge-driven mitigation with performance without mitigation.

4.4.5 Participants

IHMC recruited 26 participants, including students and staff from the University of West Florida community, for participation in the Pre-CVE (12) and CVE (14) tasks. All participants from this pool were naïve to the dynamics of the VE and the mitigation devices. However, they were all experienced computer game players, familiar with controlling their actions and movements within the VE with a minimum of cognitive effort. They did not have any alcoholic beverages or sedating medications (for example, cold and flu medications) for at least 12 hours prior to participation. Participants could be of any race or gender provided they met the above criteria. Pregnant women were not allowed to participate.

For the CVE, 14 males (M_{age} = 25.4 years) volunteered as participants for the experiment. Participants had an average education level of 15 years. To reduce the effect of learning for this experiment, participants were chosen who rated their skill level at playing first-person shooter games as average to above average. The average skill rating was 3.4/5 (Range = 2-4), with only one person rating himself as a 2/5. Overall, participants' average time playing was 5.7 hours per week.

4.4.6 Dependent Measures

Dependent variables were defined for each scenario. There were five types of dependent variables: quantitative, behavioral, indirect, subjective, and gauge-related. Below is a list

of dependent variables, by type. The dependent variables are listed with relevant task, and the relevant scenario is in parentheses.

- Quantitative
    - Engage Foes: hits on OPFOR (Scenario 1, 2)
    - Engage Foes: hits taken (Scenario 1, 2)
    - Engage Foes: shooting accuracy (Scenario 1, 2)
    - Navigation to Objective: total traversal time (Scenario 1, 2)
    - Navigation Through Unfamiliar Area: total traversal time to safe zone (Scenario 2)
    - Navigation Through Unfamiliar Area: avoid ambush zones (Scenario 2)
    - Coordinate Medevac: total time to complete procedure (Scenario 2)
    - Identify Targets: detection accuracy (Scenario 3)
- Behavioral
    - Manage Communications: message comprehension via expected observable overt behavior (Scenario 1)
    - Manage Communications: message comprehension via response with information (Scenario 1, 2)
    - Coordinate Medevac: correct responses to medevac questions (Scenario 2)
- Indirect
    - Manage Communications: situation awareness post-trial questions (Scenario 1, 2)
- Subjective
    - Workload: NASA TLX (Scenario 1, 2)
    - Preferences (Scenario 1, 2)

4.4.7 Experiment Protocol

The participants received each of the first two scenarios in one of the mitigation strategy conditions before transitioning to the second condition, repeating Scenarios 1 and 2. Thus, for Scenarios 1 and 2, this evaluation was a within-participants design, as each participant saw both scenarios in both the mitigated and unmitigated cases. Scenario 3 was presented as the final trial. Half of the participants saw Scenario 3 with mitigation, and half of the participants saw it without the mitigation. Thus, for Scenario 3, the experiment is a between-participants design. The experiment trial sequence is illustrated in Figure 27.

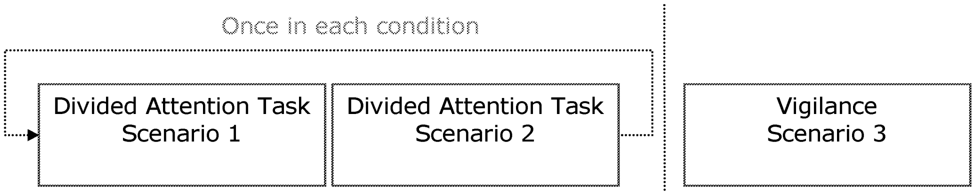


Figure 27. Order of the three experiment scenarios.

The CVE had 14 participants. Table 13 shows the experiment design for each of the participants.

Table 13. Experiment design.

		Scenarios (within-participants)		Scenario (between-participants)
		1	2	3
Augmentation Condition	Off	Unmitigated	Unmitigated	Unmitigated
	On	Mitigated	Mitigated	Mitigated

Table 14 illustrates the participant counterbalancing.

Table 14. Participant counterbalancing.

Scenarios		1A	2A	1B	2B	Vigilance
Participants	s1	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s2	mitigated	mitigated	unmitigated	unmitigated	mitigated
	s3	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s4	mitigated	mitigated	unmitigated	unmitigated	mitigated
	s5	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s6	mitigated	mitigated	unmitigated	unmitigated	mitigated
	s7	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s8	mitigated	mitigated	unmitigated	unmitigated	mitigated
	s9	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s10	mitigated	mitigated	unmitigated	unmitigated	mitigated
	s11	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s12	mitigated	mitigated	unmitigated	unmitigated	mitigated
	s13	unmitigated	unmitigated	mitigated	mitigated	unmitigated
	s14	mitigated	mitigated	unmitigated	unmitigated	mitigated

4.5 Phase 2b IHMC Results

Data were collected at the IHMC facility in Pensacola, Florida, between June 25 and July 6, 2004. This section details the analyses done on the performance data collected at the IHMC CVE.

4.5.1 Scenario 1: Multitasking

This scenario focused on the divided attention bottleneck in multitasking and consisted of the participant performing three tasks. The mitigation strategy employed in this scenario was the Communications Scheduler. Table 15 details each task, the mitigation (if applicable), the metrics associated with that task, and the performance improvement goal.

Table 15. Task metrics for Scenario 1

Task	Metric	Mitigation	Goal
Manage Communications	• Message Comprehension • Situation Awareness	Communications Scheduler	Improvement
Navigate to Objective	• Runtime	None	No Decrement
Engage Foes	• Hits Taken • Hits on OPFOR • Shooting Accuracy	None	No Decrement

Message Comprehension: Participants in the unmitigated condition correctly responded to 57 of 143 possible messages (39.9%). Participants in the mitigated condition correctly responded to 114 of 143 messages (79.7%). The mitigated condition shows a significant ($p < 0.0001$) performance increase of 100%, as shown in Figure 28.

Situation Awareness: Participants in the unmitigated condition correctly responded to 22 of 84 SA questions (26.2%). Participants in the mitigated condition correctly responded to 49 of 84 SA questions (58.9%). The mitigated condition shows a significant ($p = 0.009$) performance increase of 125%, as shown in Figure 28. SA was key to the ability to effectively manage mission priorities and coordinate with team members. Performance in this area was particularly difficult in high-workload periods, as evidenced by the low overall scores. Even with the dramatic improvement as a result of the mitigation strategy, there is an opportunity here for further improvement.

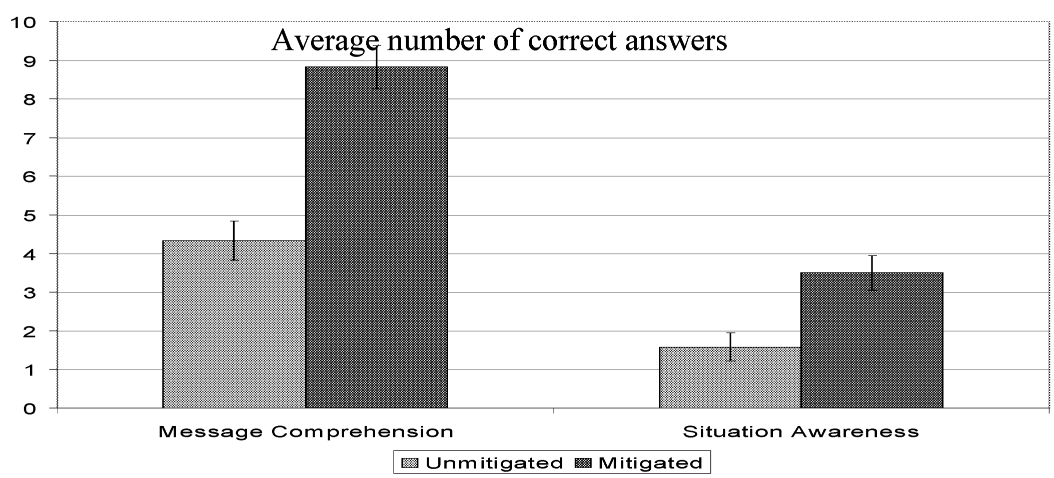


Figure 28. Communications management task metrics.

Runtime: The total runtime (seconds) for the mitigated condition (M = 965 seconds) was significantly longer than the runtime for the unmitigated condition (M = 469 seconds),

$t(12) = -8.29$, $p < 0.001$. This result, illustrated in Figure 29, is in the expected direction. In the unmitigated condition, participants often were not able to attend to the messages ordering them to hold at specific locations due to high workload. In the mitigated condition, the Communications Scheduler presented these high-priority messages with a visual and auditory cue while shifting lower priority messages to the Tablet PC during high workload so that participants had a better chance of hearing and comprehending the messages. Therefore, more participants heard the hold messages to avoid ambushes, which increased the time it took to complete their mission.

Hits Taken: There was no significant performance change for hits taken while engaging foes ($p = 0.29$), as illustrated in Figure 29.

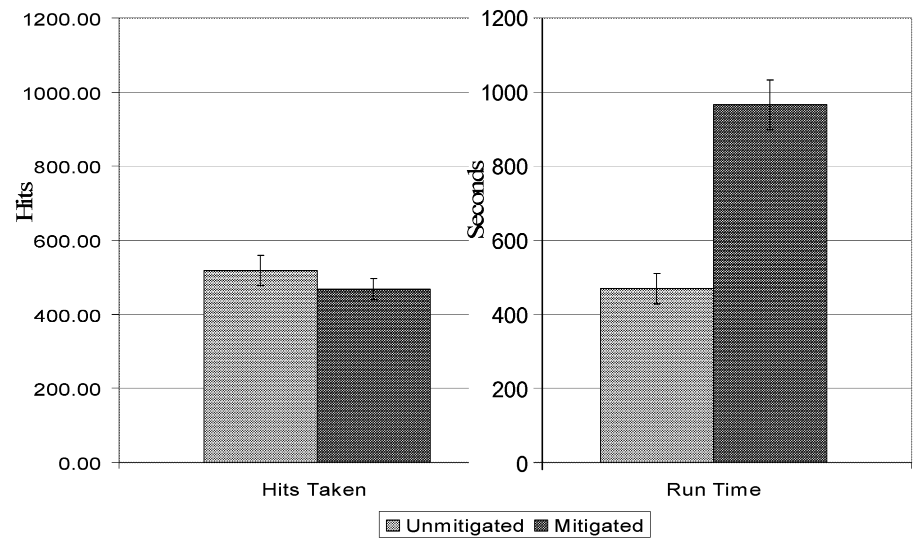


Figure 29. Scenario 1 metrics: Hits taken and runtime.

Hits on OPFOR: There was a significant difference between the unmitigated (M = 518 seconds) versus the mitigated (M = 468) for how many times the participant was able to shoot OPFORs, $t(12) = 2.93$, $p = 0.013$. Participants shot the OPFOR significantly fewer times during the mitigated condition, as shown in Figure 30. This result was expected because participants held more often when instructed to and avoided ambushes where they would encounter more enemy Soldiers. Overall, the mitigation allowed participants to have fewer encounters with enemy Soldiers, resulting in fewer chances to shoot at and hit OPFOR.

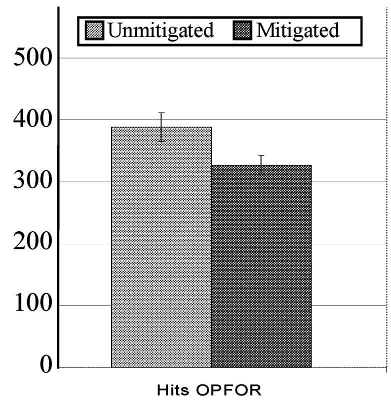


Figure 30. Scenario 1 metrics: Number of times participant hit OPFOR.

Shooting Accuracy: There was no significant performance change for shooting accuracy while engaging foes (p = 0.06), as shown in Figure 31.

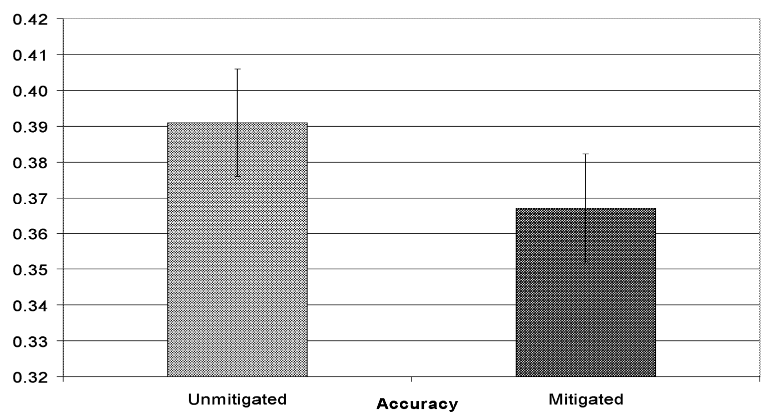


Figure 31. Scenario 1 metrics: Shooting accuracy.

4.5.2 Scenario 2: Multitasking with Return to Safe Zone and Medevac Tasks

This scenario focused on the divided attention bottleneck in multitasking and consisted of the participant performing two tasks. The mitigation strategies employed in this scenario were the Tactile Navigation Cueing System and the Medevac Negotiation Tool. Table 16 details each task, the mitigation (if applicable), the metrics associated with that task, and the performance improvement goal.

Table 16. Task metrics for Scenario 2.

Task	Metric	Mitigation	Goal
Navigate thru Unfamiliar Area	• Time to Safe Zone • Ambushes Encountered	Tactile Navigation Cueing	Improvement
Coordinate Medevac	• Questions Answered • Time to complete	Medevac Negotiation Tool	Improvement
Engage Foes	• Hits Taken • Hits on OPFOR	None	No Decrement

Time to Safe Zone: The total time it took a participant to reach the safe zone was greater for unmitigated participants than mitigated participants. The data is illustrated in Figure 32. The average performance improvement was 20%, but this difference was not significant (p = 0.30).

Ambushes Encountered: Participants in the unmitigated case were almost four times as likely to navigate into an ambush as participants in the mitigated case. Unmitigated participants (N = 12) ran into 19 ambushes, while the mitigated participants (N = 12) ran into five ambushes. The difference was significant ($p<0.003$). The mitigation resulted in a 380% performance improvement.

Hits Taken: Overall, participants took fewer hits from the OPFORs during the mitigated (M = 102) versus the unmitigated (130) condition, $t(12) = 2.42$, $p = 0.03$. The data is illustrated in Figure 32. This result was as expected. Participants in the mitigated condition potentially encountered fewer ambushes because their ability to navigate the safe route back to the safe zone was improved by the presence of the tactor and visual cues. This resulted in participants seeing fewer enemy forces and thus receiving fewer hits.

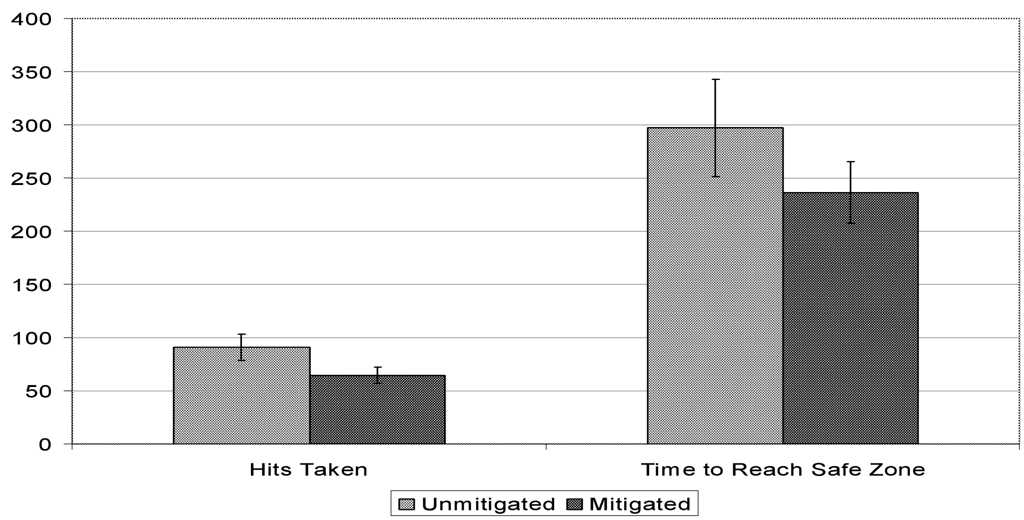


Figure 32. Scenario 2 metrics: Hits taken and time to reach safe zone.

Medevac Questions Answered: Participants in the unmitigated case answered 50 questions correctly out of a possible 98 (51% correct). Participants in the mitigated case

answered 98 of 98 questions correctly (100% correct). Medevac messages were of high priority to mission success, and the mitigations enabled the participants to appropriately attend to their content and respond accordingly. Thus, the mitigation was able to significantly ($p = 0.004$) increase performance by 95%, as shown in Figure 33.

Time to Complete Medevac: Participants were able complete the medevac task significantly faster in the mitigated condition ($p < 0.001$), resulting in a 303% performance improvement, as shown in Figure 33.

Hits on OPFOR: Participants had fewer opportunities to engage the OPFOR and therefore had fewer hits on OPFOR in the mitigated ($M = 22$) versus the unmitigated ($M = 27$) condition, $t(12) = 2.47$, $p = 0.029$. This was expected, since participants in the mitigated condition potentially encountered fewer ambushes because their ability to navigate the route back to the safe zone was improved by the presence of the tactor and visual cues. This resulted in participants seeing fewer enemy forces and thus having fewer opportunities to hit the OPFOR.

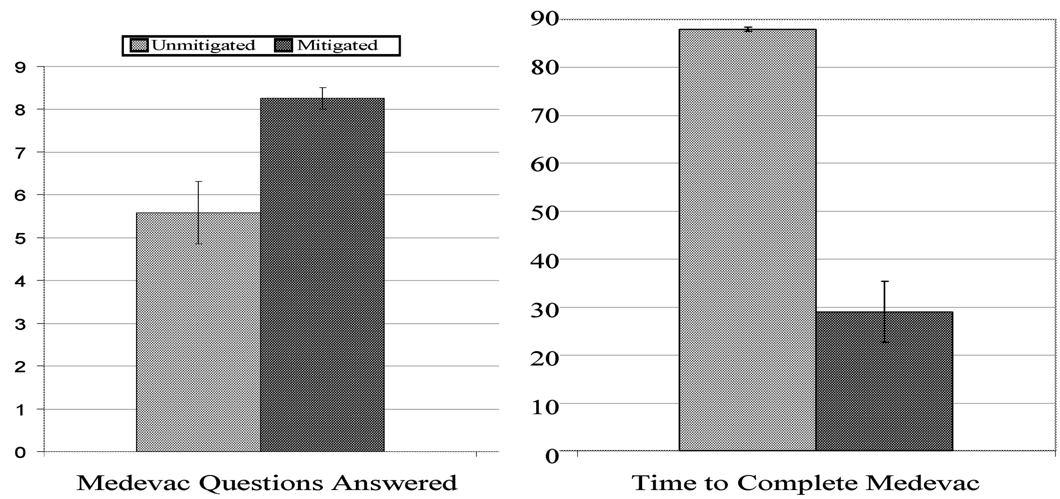


Figure 33. Scenario 2 metrics: Medevac questions answered and time to complete medevac.

4.5.3 Scenario 3: Vigilance Monitoring Task

This scenario focused on the sustained attention bottleneck in a vigilance paradigm, where participants spent long durations monitoring in a target-deficient area. The participants performed the tasks of target detection and target identification. The mitigation strategy employed in this scenario was the Mixed-Initiative Target Identification System. Table 17 details each task, the mitigation (if applicable), the metrics associated with that task, and the performance improvement goal.

Table 17. Task metrics for Scenario 3.

Task	Metric	Mitigation	Goal
Target Identification	• Accuracy of Target ID	Mixed-Initiative Target Identification System	Improvement

Target Identification: Recall that the vigilance scenario had three stages. Stage 1 (the first 5 minutes) was considered the baseline condition of alert human performance. No mitigation was employed in stage 1. On average, participants had a baseline performance in stage 1 of 65.8%. Stage 2 consisted of a 20-minute interval designed to induce a vigilance decrement. Stage 3 was the final 5-minute period, where some participants performed the task with mitigation, which was set at an accuracy rating of 68%. Unmitigated participants in stage 3 had an accuracy of 66.2%. Thus, on average, the experiment was not able to produce the decremented human performance desired in a vigilance experiment. Nonetheless, participants in the mitigated condition performed at an accuracy of 85%, much better than the human (66.2%) or automation (68%) accuracy alone. The 30% performance improvement, shown in Figure 34, was significant ($p = 0.022$).

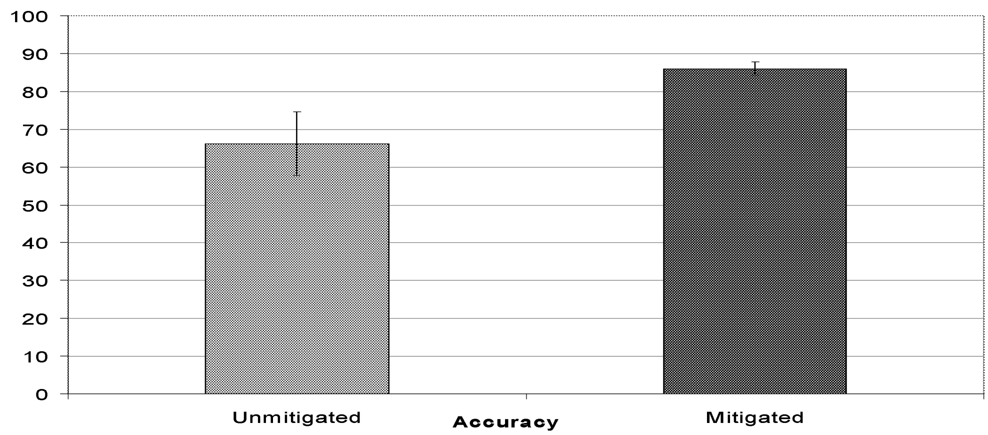


Figure 34. Scenario 3 metric: Target identification accuracy.

4.5.4 Subjective Results

The goal of the mitigations was to improve performance while not having detrimental effects on workload. Table 18 lists, for the unmitigated participants versus the mitigated participants in Scenario 1, the average TLX workload score for each rating scale in the NASA TLX, in addition to the total workload average. None of the scales nor the total overall workload was significantly different.

Table 18. Workload ratings for Scenario 1.

Scenario 1	Mental	Physical	Temporal	Performance	Effort	Frustration	Total
Unmitigated	7.07	3.39	5.75	3.25	7.00	6.75	5.54
Mitigated	7.64	2.89	6.18	3.64	7.61	6.50	5.75
p-value	**0.4**	**0.24**	**0.39**	**0.53**	**0.11**	**0.61**	**0.46**

Likewise, for Scenario 2, there was no significant difference in any of the workload rating scales or in the total workload rating, as shown in Table 19.

Table 19. Workload ratings for Scenario 2.

Scenario 2	Mental	Physical	Temporal	Performance	Effort	Frustration	Total
Unmitigated	7.15	3.00	5.85	4.15	7.62	6.00	5.98
Mitigated	7.27	3.38	6.73	5.15	7.32	5.77	5.94
p-value	**0.8**	**0.1**	**0.25**	**0.06**	**0.33**	**0.75**	**0.91**

Participants were asked their preferences with regard to tasks in the scenario. Overall, 76.9% of the participants thought it was easier to perform tasks in the mitigated condition. Participants also found the mitigated condition easier for fighting (61.5%), communicating (84.6%), and navigating (76.9%). The results are summarized in Table 20.

Table 20. Participant preferences with regard to tasks in the environment.

Question	Mitigated	Unmitigated	Same	Mitigated	Normal	Same
"Fighting was easier in..."	8	3	2	61.5%	23.1%	15.4%
"Communicating was easier in…"	11	2	0	84.6%	15.4%	0.0%
"Navigation was easier in…"	10	2	1	76.9%	15.4%	7.7%
"Overall I found it easier to perform tasks in…"	10	2	1	76.9%	15.4%	7.7%

4.5.5 Bottleneck Mitigation Findings Summary: IHMC CVE

The *Communications Scheduler* mitigation of the divided attention bottleneck resulted in the following performance improvements:

- 100% improvement in message comprehension
- 125% improvement in SA
- No negative effect on ability to engage foes
- No negative effect on workload
- 85% of participants felt communication easier with augmentation

The *Tactile Navigation Cueing mitigation* of the sensory input bottleneck resulted in the following performance improvements:

- 20% decrease in evacuation time
- 380% decrease in number of ambushes encountered
- No negative effect on ability to engage foes
- No negative effect on workload
- 80% of participants felt navigation was easier with augmentation

The *Medevac Negotiation Assistance mitigation* of the executive function bottleneck resulted in the following performance improvements:

- 96% improvement in communication of critical information
- 303% improvement in time to complete negotiation
- No negative effect on ability to engage foes
- No negative effect on workload

The *Mixed-Initiative Target Identification mitigation* of the sustained attention bottleneck resulted in the following performance improvements:

- 30% improvement of joint human-machine performance over decremented human performance

4.6 CMU CVE System Design and Architecture

4.6.1 Component Overview

The system used at CMU was based on the same architecture used at IHMC, described in Section 4.3. The principal difference was the number of gauges integrated in the system. For the CMU CVE, the Arousal Meter, Engagement Index, and XLI were used to estimate the participants' cognitive load.

4.6.2 Conceptual System Architecture and Rationale: CMU CVE

The system architecture and rationale were very similar to the IHMC system and rationale described above; however, the CMU setup differs in the following ways:

- The scenarios at CMU required the same multitasking inherent to dismounted operations; however, at CMU the participant played the role of a rooftop lookout who was monitoring a more well-defined space—4 x 4 array of windows in adjacent building. Like the IHMC scenario, CMU participants were also required to monitor radio communication and shoot at enemies.
- Participants immersed in a Panda 3d-based VE with motion-capture and tracking instead of the modified desktop simulation VE (Quake).
- Subset of gauges run: XLI, Engagement Index, and Arousal Meter:
    - o Did not use Stress Gauge, since its most valuable input, pupilometry, could not be attained due to limitations imposed by virtual reality head-mounted display and availability of head-space for mounting eye tracker.
    - o Did not use P300 due to basic technical constraint of injecting precisely timed event triggers directly into EEG signal.
- Mitigation trigger included a rate of change threshold in addition to numerical threshold—to increase sensitivity to early stages of task load increase before gauge values exceeded numerical threshold.
- Employed only a single mitigation strategy at CMU—gauge-enabled scheduling of messages.
- Single scenario with dual tasks within it.

4.6.3 Mitigation Strategies and Rationale

The premise of the mitigation strategy was to intelligently schedule incoming radio messages based on the current task load of the participant. For example, if participants were actively monitoring a building and maintaining their counts, any additional incoming messages would create a catastrophic interference that degraded the memory for all counts; however, if the gauges detected a high task load and deferred delivery of auditory messages until the participant was under a low task load or reported his/her status, this would reduce the likelihood of a counting-task interference within this paradigm. Under perfect mitigation, participants would maintain only their three egocentric counts (total number of friendlies encountered, total number of enemies encountered, total shots fired) during the first monitoring part of the scenario; once they reported their status, the deferred radio messages would be presented, only requiring the participants to maintain two counts (total of reported number of friendlies reported by Squad A, total number of enemies reported by Squad A) before reporting them out. Perfectly applied mitigation would reduce the occurrence of counting interference and reduce from five to three the number of counts to be maintained during the high-workload monitoring phase. See Table 21 for an overview.

Table 21. Dual task pair.

Dual Task Pair		
	Primary	**Secondary**
Subtasks	•Building monitoring •Shooting enemies •Radio monitoring •Maintaining 3-5 counts	•Radio monitoring •Maintaining 2 counts
Report at end of ~ 1-minute period	3-5 counts	2 counts
Ideal mitigation response	Radio messages deferred – only has to maintain 3 counts	2 counts still (though radio communications are more frequent)

The role of the Mitigation Agent at CMU was to enable radio message scheduling based on the sensed cognitive state of participants. It was configured to identify when participants were experiencing the high workload associated with the multitasking load inherent to actively monitoring the building while maintaining multiple counts in working memory. When the Mitigation Agent identified such a state, it deferred incoming radio messages to be played during the lower workload in the secondary task period; otherwise, if high workload was not identified, the Mitigation Agent allowed radio messages to pass through to the participants.

The Mitigation Agent assumed it will receive input from the Z-norm Engagement, XLI, and Arousal Meter gauges. It used simplified logic that determined if the system is in one of eight possible states:

- All three gauges up and running (all have certainty > = 0)
- At least two gauges up and running (Arousal and Z-norm have certainty > = 0)
- At least two gauges up and running (Arousal and XLI have certainty > = 0)

- At least two gauges up and running (XLI and Z-norm have certainty > = 0)
- Only one gauge running (Arousal has certainty > = 0)
- Only one gauge running (Z-norm has certainty > = 0)
- Only one gauge running (XLI has certainty > = 0)
- No gauges up and running

Thresholds were selected for all gauges to maximize the differences between hits (high workload during primary task) and false alarms (high workload during secondary task) for the gauge-validation study data. Gauge thresholds considered numerical thresholds as well as recent rate of change (ROC) for each gauge individually. Gauges were considered high if either the numerical or ROC threshold was met. The gauge would return a Boolean value to turn on mitigation if the rules were satisfied and turn it off if they were not. The mitigation triggering rule set logic is shown in Figure 35.

1. If (2 of the 3 is TRUE):
    - Arousal Meter is > .25 OR increased by at least .35 over last 5 seconds
    - Z-norm is > - 1.5 OR increased by at least .25 over last 5 seconds
    - XLI ROC over 3 samples (~ 6 sec) < 0

    THEN Mitigate ON, ELSE Mitigate OFF
2. IF((Arousal Meter is > .25 OR increased by at least .35 over last 5 seconds) OR (IF (Z-norm is > - 1.5 OR increased by at least .25 over last 5 seconds)), THEN Mitigate ON, ELSE Mitigate OFF
3. IF((Arousal Meter is > .25 OR increased by at least .35 over last 5 seconds) (XLI ROC over 3 samples (~ 6 sec) < 0), THEN Mitigate ON, ELSE Mitigate OFF
4. IF((XLI ROC over 3 samples (~ 6 sec) < 0) OR (Z-norm is > - 1.5 OR increased by at least .25 over last 5 seconds)), THEN Mitigate ON, ELSE Mitigate OFF
5. IF (Arousal Meter is > .25 OR increased by at least .35 over last 5 seconds), THEN Mitigate ON, ELSE Mitigate OFF
6. IF (Z-norm is > - 1.5 OR increased by at least .25 over last 5 seconds), THEN Mitigate ON, ELSE Mitigate OFF
7. IF (XLI ROC over 3 samples (~ 6 sec) < 0), THEN Mitigate ON, ELSE Mitigate OFF
8. Mitigate OFF

Figure 35. Mitigation trigger rule set logic for the CMU CVE.

4.7 Phase 2b CMU Concept Validation Experiment

4.7.1 Experiment Objectives

Dismounted Soldiers are required to maintain counts of their possessions, such as ammunition, as well as encountered entities, such as civilians and combatants. Several factors inherent to dismounted operations conspire to negatively affect Soldiers' ability to maintain items in working memory. These include:

- During high-paced operations, Soldiers do not have the opportunity to update and rehearse their respective counts, which results in failure to maintain an accurate count;
- The inherently stressful nature of dismounted operations consistently disrupts Soldiers' capacity to maintain items in working memory;
- Frequent task switching and communications saturation interferes with maintaining accurate counts.

The objective of the CMU CVE was to demonstrate dramatically improved performance of an operationally relevant task using gauge-driven scheduling on a mobile participant.

The basic approach was to extend the application of several derived gauges that use the raw input from EEG and ECG systems in this task-scheduling context. A rule-based logic that reasoned about the current state was used, as well as the direction and rate of change of three gauges (Arousal Meter, XLI, and Engagement Index) in order to determine whether incoming messages should be deferred until a later time when the message in question would be less likely to interfere with ongoing task requirements. Expectations included answering the following questions:

- Can a task-relevant cognitive state be reliably detected in a mobile participant?
- Can information about cognitive state be used to schedule incoming messages?
- Will the gauge-enabled scheduling produce performance improvements?

4.7.2 Operational Scenario

The participant was asked to play the part of a military lookout on a virtual rooftop in a simplified urban environment. He or she wore a lightweight, motion-tracked head-mounted display and was given a motion-tracked M16 rifle prop. The gun prop was visible in the VE and produced a red laser dot on objects, indicating precisely where the gun was being aimed. In the environment, the participant was surrounded by four buildings, each in one of the cardinal directions: north, south, east, or west. Each building had four columns of evenly spaced windows. The windows of the top four floors on each of the buildings were open, producing a four-by-four array of windows past which friendly or enemy Soldiers would walk.

Computer speakers in the room allowed for simulated radio broadcasts to be heard by the participant. At the beginning of each section of a given trial, a radio message was played instructing the participant to face a particular direction. After that message, groups of friendly and enemy Soldiers walked past various windows in the building. Radio messages were also being broadcast periodically, giving numbers of friendly or enemy Soldiers spotted by other lookouts. Each message ended with the name of the team leader that it is addressed to (e.g., Bravo leader). The participant was instructed to do the following things:

1. Shoot as many enemy Soldiers as possible
2. Keep a running count of the number of friendly Soldiers seen
3. Keep a running count of the number of enemy Soldiers seen
4. Keep a running count of the number of bullets fired
5. Keep a running count of the number of friendly Soldiers reported over the radio, only taking into account messages addressed to the Bravo team leader
6. Keep a running count of the number of enemy Soldiers reported over the radio, only taking into account messages addressed to the Bravo team leader

At prescribed times, a radio message was given to “Report your status.” At this time, the participant verbally reported the running counts that he or she had been keeping. After

reporting was completed, the participant's counts were reset to zero and the experiment continued.

Each trial was divided into several repeated blocks, each consisting of two parts. In the first part, groups of Soldiers walked past the windows of the building that the participant was facing. The Soldiers came in groups, appearing one right after another. Friendly and enemy reports were sent over the radio during this period, in addition to visually identifying and engaging the Soldiers in the windows. In the second part of the block, no friendly or enemy Soldiers walked past any windows, and the participant was only required to deal with radio messages that he or she received. A "Report your status" message was given at the end of each of these parts.

Figure 36 depicts the task environment containing two enemies (green) and three friendlies (tan).

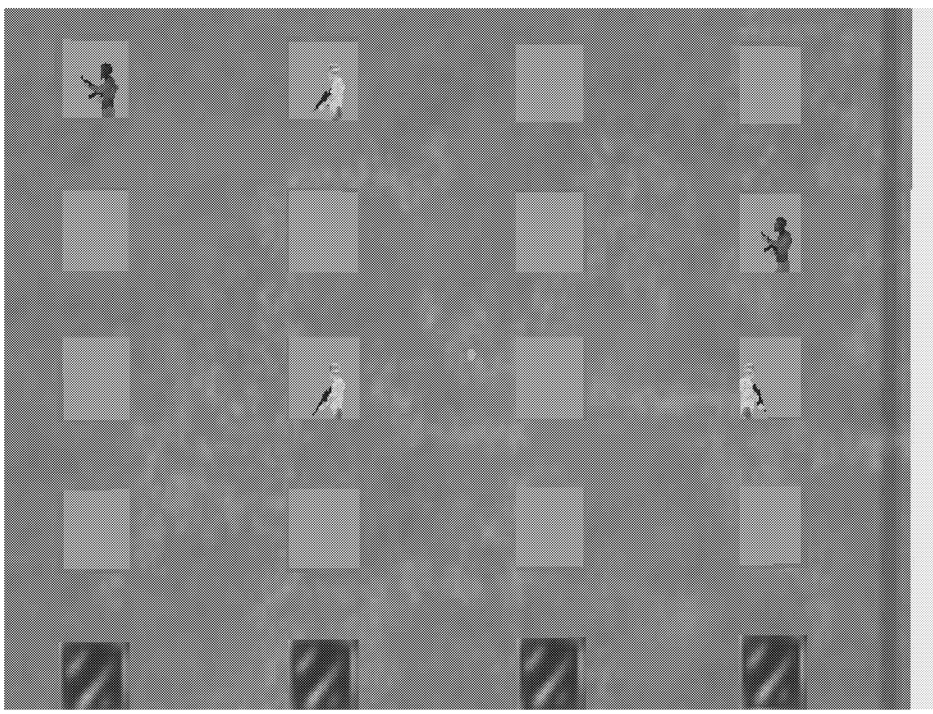


Figure 36. The CMU CVE environment.

4.7.3 Experiment Hypothesis

Within this paradigm, the critical comparison was between counting performance for the gauge-driven scheduling condition and the randomly scheduled unmitigated condition. If the gauge logic was able to detect both high and low task load, the mitigation would defer messages from the high-load period until the low-load period. The primary hypothesis was that gauge-enabled scheduling of incoming radio reports would dramatically reduce the catastrophic interference of managing; thereby, participants would have dramatically fewer reporting errors with regard to the counts that they had to maintain. This was based on the behavioral data collected on a similar task in the same environment, where a 300% performance improvement was found (see Dorneich et al., 2004b).

4.7.4 Experiment Design

The design was a 2 (augmentation: mitigated and unmitigated) x 4 (pairs of primary and secondary tasks) factor experiment. Each participant completed a set of four pairs of primary-secondary tasks in one augmentation condition (i.e., mitigated or unmitigated) and then completed the second set in the other augmentation condition. The order in which the participants received the augmentation was counterbalanced.

The main hypothesis was that gauge-enabled scheduling would produce dramatic performance improvements during the primary task periods. A secondary hypothesis was that gauge measures would reliably detect relevant cognitive states defined as high task load during the dual monitoring of the primary task periods.

Primary Task Period:

- Monitored building for enemies and friendlies
- Shot enemies
- Monitored radio communications
- Maintained cumulative counts of:
    - Number of friendly Soldiers seen
    - Number of enemy Soldiers seen
    - Number of bullets fired
    - Number of friendly Soldiers reported over the radio, only taking into account messages addressed to the Bravo team leader
    - Number of enemy Soldiers reported over the radio, only taking into account messages addressed to the Bravo team leader
- Reported counts at the end of the primary task period

Secondary Task Period:

- Monitored radio communications
- Maintained cumulative counts of:
    - Number of friendly Soldiers reported over the radio, only taking into account messages addressed to the Bravo team leader
    - Number of enemy Soldiers reported over the radio, only taking into account messages addressed to the Bravo team leader
- Reported counts at the end of the secondary task period

Participant performance was compared under an Augmentation ON condition (gauge-based scheduling) with a random scheduling condition. With Augmentation ON, gauge values were used to determine if the cognitive state of the participant was overloaded to a point where the system should defer radio messages during the primary task period. Under the random scheduling condition, radio messages were randomly presented within

the primary-secondary task pair period. Table 22 represents the research design of a 2 (mitigation) x 2 (block) within-participants design counterbalanced for order of presentation.

Table 22. Experiment design of CMU evaluation.

Part. group	Block 1				Block 2			
A	Mitigation ON	Mitigation ON	Mitigation ON	Mitigation ON	Mitigation OFF (Random)	Mitigation OFF (Random)	Mitigation OFF (Random)	Mitigation OFF (Random)
	Primary-Secondary Task Pair A	Primary-Secondary Task Pair B	Primary-Secondary Task Pair C	Primary-Secondary Task Pair D	Primary-Secondary Task Pair A	Primary-Secondary Task Pair B	Primary-Secondary Task Pair C	Primary-Secondary Task Pair D
B	Mitigation OFF (Random)	Mitigation OFF (Random)	Mitigation OFF (Random)	Mitigation OFF (Random)	Mitigation ON	Mitigation ON	Mitigation ON	Mitigation ON
	Primary-Secondary Task Pair A	Primary-Secondary Task Pair B	Primary-Secondary Task Pair C	Primary-Secondary Task Pair D	Primary-Secondary Task Pair A	Primary-Secondary Task Pair B	Primary-Secondary Task Pair C	Primary-Secondary Task Pair D

4.7.5 Participants

Recruited participants included students and staff from the CMU community. All participants from this pool were not necessarily naive to the dynamics of the dismounted Soldier simulation and the control input devices. They had no alcoholic beverages or sedating medications (for example, cold and flu medications) for at least 12 hours prior to participation. Participants could be from any race or gender, provided they met the above criteria. Pregnant participants were acceptable.

4.7.6 Dependent Measures

Within the CMU task environment, the metrics of success (MOS) were performance on the counting tasks that required participants to attend to the building they were monitoring, as well as incoming radio reports for which they maintained the updated counts in working memory. The counts maintained were the total number of friendlies encountered, the total number of enemies encountered, total shots fired, total of reported number of friendlies reported by Squad A, and total number of enemies reported by Squad A. Participants received radio reports of encounters at random intervals from multiple squads. They were instructed to only pay attention to reports from Squad B and to ignore reports from Squad A. At the end of each experiment block, participants were instructed to report all counts. Participants received a score that is a function of discrepancy between reported count and actual count. Participants received one point for an accurate count, with a maximum score of five for each experiment block.

Dependent variables:

1. Count performance measured by a function of the discrepancy between reported count and actual count
2. Shooting accuracy (number of shots fired/ enemies hit)
3. Discrimination between friends and foes
4. Number of messages deferred during each phase of the experiment

5. Neurophysiological and physiological changes measured by:
    a. Engagement Index (EEG based)
    b. XLI (EEG based)
    c. Autonomic Arousal (ECG based)
6. Post-trial subjective report of cognitive workload, including NASA TLX subscales

4.7.7 Experiment Protocol

Table 23 describes the schedule that was maintained for the experiment procedure.

Table 23. Experiment protocol for Phase 2b CMU CVE.

Protocol	Time Est.
Overview of Experiment	**15 minutes**
General instructions	
Consent form	
Demographics form	
Experimenter will exclude participant based on:	
Non-native English speaker	
Red-green colorblindness	
Excluded participants	
Experimenter places ECG and EEG on participant	**30 minutes**
Participant enters virtual environment	**15 minutes**
Interactive training regimen	
Experiment Trials	**45 minutes**
Participant completes Block 1	
Completing NASA-TLX workload scale	
Break	
Participant completes Block 2	
Completing NASA-TLX workload scale	
Debriefing	**5 minutes**
Experimenter answers any questions	
Participant receives $20 remuneration	

4.7.8 Data Analysis Methodology

Data analysis started with 2 (condition: mitigated, random) x 2 (task: primary, secondary) ANOVAs (Analysis of Variance) looking for main effects and interaction for gauge measures and reported count accuracy. Significant interactions were followed up with comparisons. In addition, t-tests were performed to compare the effect of condition (mitigated, random) on shooting performance (hit rate) and perceived workload (NASA TLX), as well comparing primary and secondary tasks for XLI. Accordingly, all reported p values are for either the ANOVA or t-test for the respective measure.

The metric of success for the CMU CVE was the percentage of reported counts correct following the primary task period. Performance improvement was calculated as follows:

- (Mitigated % correct – Random % correct)/Random % correct for each participant
- Average improvement over ten participants

Analyses were performed on the following measures:

- Reported count accuracy
  - Total error rate (delta from accurate count)
- Identifying and shooting enemies
  - Hit rate: enemies shot/number of enemies encountered
- Subjective workload (NASA TLX)
  - Overall Workload
  - Mental Demand, Physical Demand, Temporal Demand
  - Performance, Effort, Frustration
- Gauge state comparing conditions
  - Engagement Index (value, ROC)
  - Arousal Meter (value, ROC)
  - XLI (ROC)
- Performance improvement
  - Primary task counting performance
  - (Mitigated performance – Random performance)/Random performance.

4.8 Phase 2b CMU Results

4.8.1 Reported Count Accuracy

This measure captured the absolute value of the discrepancy (error) between counts reported by the participant and the actual counts. The implication was that the greater discrepancy reflected overall poorer counting performance. As expected, significantly more errors were found in primary task completion compared with secondary task completion across both mitigation conditions—since participants were required to maintain at least three counts while performing a coincident building monitoring task.

To evaluate the impact of the gauge-based mitigation, the mitigated and random condition performances were compared during the critical primary task period, as shown in Figure 37. As expected, primary-mitigated condition showed marginally significant fewer count errors than primary-random condition ($p < .009$); see Table 24 and Table 25. This comparison was related to the performance improvement metric, since it reflects the relative benefit of mitigated counting performance over random counting performance.

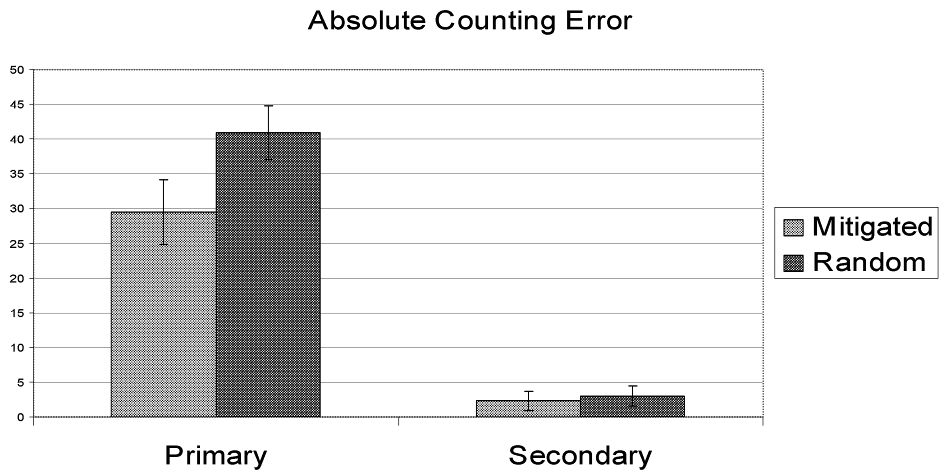


Figure 37. Absolute counting error.

Table 24. ANOVA for absolute counting error.

Results				
Analysis	df	F-value	p-value	Significant
Condition	1	9.35	0.01400	Yes
Task	1	55.59	0.00004	Yes
Condition x Task	1	9.83	0.01202	Yes

Table 25. Comparisons.

Follow-ups

Comparison	t	df	p-value	significance
Mitigated-Primary - Mitigated-Secondary	5.79	9	0.00026342	Yes
Mitigated-Primary – Random-Primary	–3.3	9	0.00919264	Marginal
Mitigated-Primary – Random-Secondary	5.16	9	0.00059401	Yes
Mitigated-Secondary – Random-Primary	–8.72	9	0.0000110	Yes
Mitigated-Secondary – Random-Secondary	–0.53	9	0.60792697	No
Random-Primary – Random-Secondary	8.11	9	0.000020	Yes
Alpha = 0.05/6 = 0.008				

4.8.2 Identifying and Shooting Enemies (Hit Rate)

For this measure, the goal was to confirm that the mitigation strategy at a minimum "did no harm" to this ancillary task performance. Hit rate was calculated by dividing the number of enemy hits by the number of enemies appearing in the monitored building. There was effectively no difference between the mitigation conditions ($p < .48$) (see Figure 38).

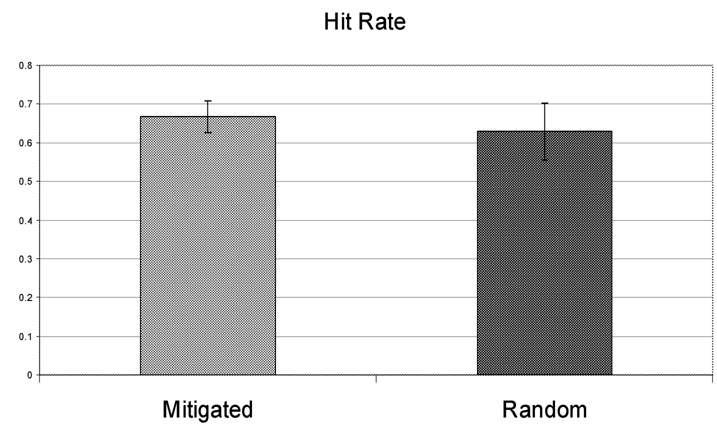


Figure 38. Hit rate.

4.8.3 Correct Counts (Performance Improvement Metric)

Performance improvement was calculated as follows: (mitigated % correct – random % correct)/random % correct. On average, the mitigated condition showed a 60% improvement over the unmitigated (random) condition, as shown in Figure 39. The difference was marginally significant ($p = 0.075$). See Table 26 for individual performance improvements.

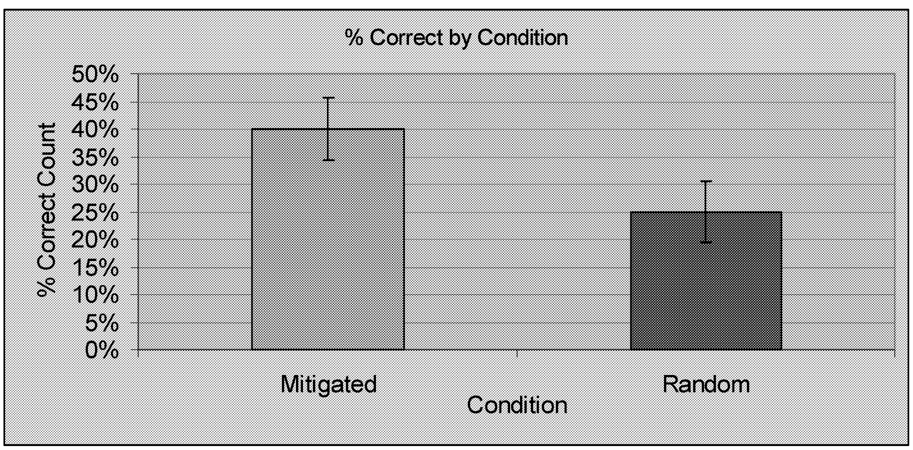


Figure 39. Average % correct count by condition.

Table 26. Average performance improvement by condition

Participant	Mitigated	Random	% Improvement
105	58.3%	16.7%	250%
106	50.0%	8.3%	500%
107	66.7%	50.0%	33%
108	8.3%	16.7%	–50%
109	33.3%	8.3%	300%
110	41.7%	25.0%	67%
111	41.7%	58.3%	–29%
112	16.7%	33.3%	–50%
113	50.0%	8.3%	500%
114	33.3%	25.0%	33%
Average	**40.0 %**	**25.0 %**	**60 %**

4.8.4 Subjective Workload (NASA TLX)

The AugCog team analyzed this data to confirm that the mitigation strategy did not negatively affect participants' perceived workload, as measured by NASA TLX subscales. Participants reported a marginally significant ($p < .06$) lower Mental Workload (see Table 27) for mitigated compared with random conditions. For all other measures (except Physical Demand), participants reported a numerically lower workload for mitigated compared with random conditions, as shown in Figure 40.

Table 27. Mental demand.

Comparison	t-value	df	p-value
Mental demand	–2.0147	14	0.06355

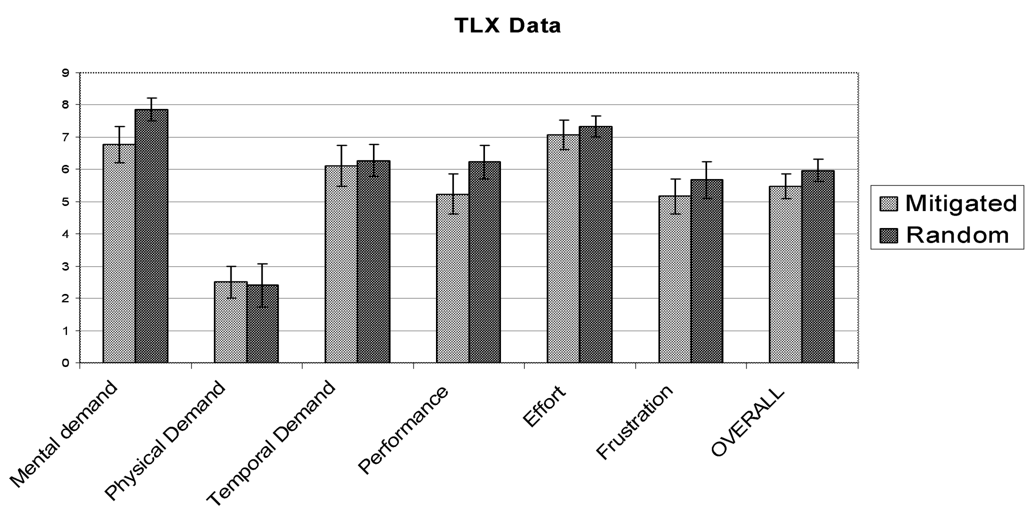


Figure 40. Workload scales for the CMU CVE participants.

4.8.5 Gauge State Comparisons

The individual gauge results were analyzed to evaluate individual gauge response to different conditions (mitigated, random) and tasks (primary, secondary) as well as to confirm expectations developed during the gauge validation studies. Differences were expected in gauge response to:

- Task condition: primary to be higher than secondary
- Critical comparison between mitigated primary (higher) vs. mitigated secondary– to confirm the thresholds used to differentiate between these factors

4.8.5.1 Engagement Index Numerical Threshold

There was a numerical difference ($p < .16$) between the primary and secondary tasks in the expected direction (see Figure 41)—higher for primary—as well as a numerical difference ($p < .16$) between the mitigated and random conditions in the expected direction—lower in mitigation (see Table 28). There was a large numerical difference (0.8) between mitigated-primary and mitigated-secondary, which provided some confirmation regarding the selected threshold; in fact, 7 of 10 participants exhibited this difference.

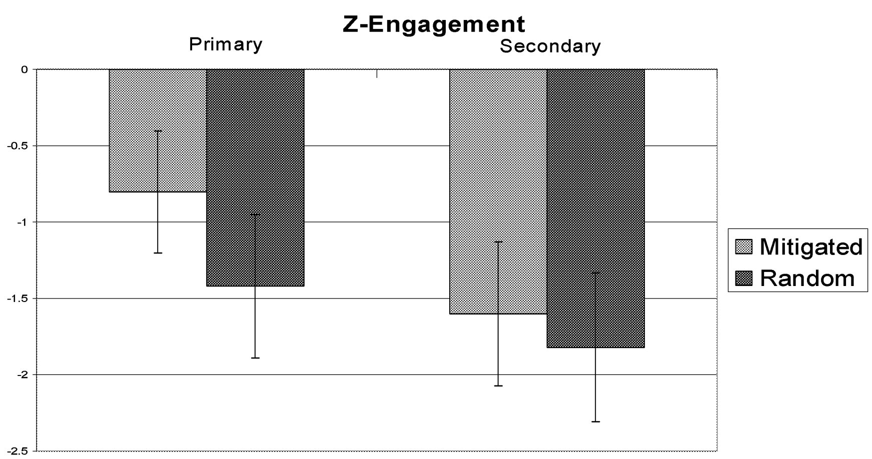


Figure 41. Z-Engagement for primary and secondary tasks.

Table 28. ANOVA for Z-Engagement gauge.

Results	df	F	p-value
Condition	1	2.32	0.16
Task	1	2.3	0.16
Condition x Task	1	1.93	0.2

4.8.5.2 *Engagement Index Rate of Change (ROC)*

There was a numerical difference ($p < .13$) between the primary and secondary tasks in the expected direction (see Figure 42 and Table 29)—higher for primary—and also a numerical difference between the mitigated-primary and mitigated-secondary conditions—thus providing some confirmation of the threshold selection; again, 7 of 10 participants exhibited this difference.

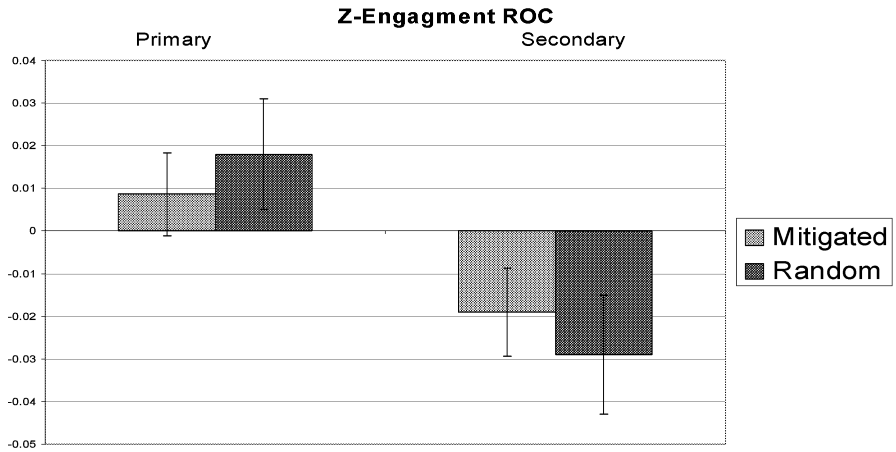


Figure 42. Z-Engagement ROC.

Table 29. ANOVA for Z-Engagement ROC.

Results	df	F	p-value
Condition	1	0.002	0.97
Task	1	2.75	0.13
Condition x Task	1	3.796	0.083

4.8.5.3 *Arousal Meter*

Due to sensor interface issues, the ECG signal quality was sufficient for analysis for only five of the participants. Given the small sample size, numerical trends were analyzed. First, the primary-mitigated condition was found to be more arousing than the secondary-mitigated condition, providing confirmation of the specific threshold selection. Second, as expected, primary task arousal was higher than secondary task arousal across both mitigation conditions, providing further confirmation of the threshold selection. Finally, as expected, the random condition was numerically more arousing than the mitigated for both primary and secondary task conditions, confirming the positive impact of the mitigation strategy. See Figure 43.

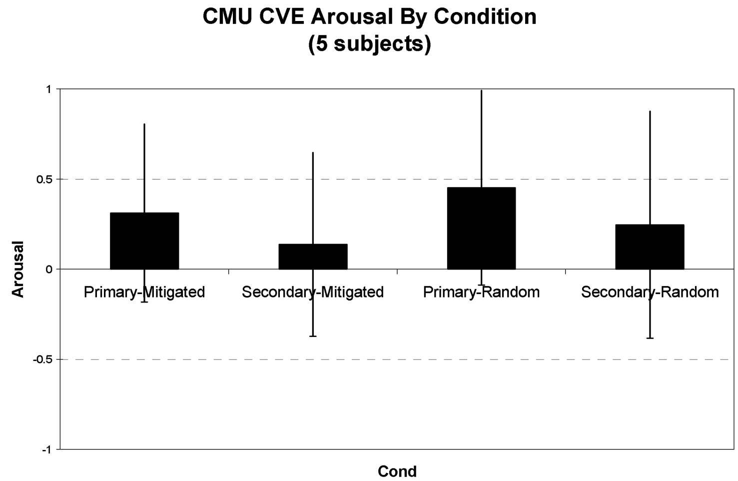


Figure 43. Arousal Meter by condition.

4.8.5.4 XLI

The differences between the primary and secondary task periods were evaluated under the random condition, revealing a numerical difference that was trending toward significance in a paired sample t-test comparing means ($p < .12$), From this, it was concluded that the same level of difference did not show up under the mitigation condition, since the manipulation effectively levels workload to reduce the difference between the primary and secondary task periods—unlike in the random condition.

In addition to this analysis, an examination was performed on the effectiveness of a simple prototype feed-forward neural network classifier in distinguishing a single discrete-time event when the participant was required to switch from a primary to a secondary task and *visa versa*. Reported here are results using a single-event neuroclassifier tuned to identify cognitive task shedding and reacquisition events across participants. The preliminary test of the neuroclassifier algorithm performed at up to 72% accuracy across both random (unmitigated) and mitigated conditions over 26 experiment sessions. Human Bionics is continuing to improve the prototype XLI neuroclassifier's accuracy in identifying discrete task switching events.

4.9 Phase 2b Discussion

The AugCog team leveraged a gauge-based mitigation strategy to produce a 60% performance improvement over an unmitigated (random) scheduling performance baseline. Honeywell substantiated the mitigation triggers by identifying numerical trends for all gauges that were consistent with the expectations and thresholds established during the gauge validation studies intended to differentiate primary and secondary tasks within this immersive virtual environment task. Moreover, the mitigation strategy produced significantly lower perceived mental workload when compared with the unmitigated condition.

Many lessons were learned, with implications for future phased development of the AugCog technologies. Some of the challenges and future development plans are outlined below.

4.9.1 System Usability Challenges

4.9.1.1 Challenges in Phase 2b CVEs

Beyond the technical challenges of building a CLIP and designing an evaluation around the system's capabilities, several setbacks were encountered simply from the participants' interaction with the prototype system, the tasks, and the virtual environment(s). Participants in both environments experienced simulator sickness due to either the time spent viewing the environment with a head-mounted display that had a restricted field of view or the sheer weight of some of the components on the participants' heads. These elements, coupled with mentally demanding tasks such as vigilance tasks requiring long periods of sustained attention, resulted in some simulator sickness and general nausea.

4.9.1.2 Future Evaluation Challenges

Fortunately, the challenges of the Phase 2b CVEs will not be encountered in field testing. However, new challenges will arise. Field testing will be concerned about sensor interaction beyond the psycho-physiological sensors that have already been integrated. Components such as accelerometers will be integrated to detect motion and head position and Global Positioning System (GPS) for location. This information will be used for context modeling to understand what task the Warfighter is engaged in and how to understand his or her cognitive state within that context. Another concern is reducing the processing requirements and sensor requirements in order to classify cognitive state.

Additional factors that will challenge these evaluations include the addition of a truly mobile individual, a limitation on the types of tasks that can be performed, and the natural component introduced with testing outdoors (terrain, daylight, temperature, winds, precipitation, etc). Plans will be made to manage these conditions, since they will be faced by the target populations.

As those challenges are addressed, others will arise as the system is integrated and tested with FFW's components. An evaluation of a system of systems will result in less control over evaluation scheduling, task definition, evaluation environment, available components for integration, and processing power. There will be challenges of using only the available communication bandwidth, computer processing power, and battery power life. Again, these conditions will need to be overcome as the technology transitions to Warfighters. Plans will be made to manage them.

4.9.2 Human-Computer Information Processing

The four key information processing bottlenecks (inputs, attention, executive function, working memory) were identified and addressed in the evaluations conducted by the Honeywell team. However, several areas were identified that cannot be overlooked if the technology is to be successfully demonstrated in an operational environment. For instance, the mitigation etiquette, or how the automation and operator interact, is a key component of the system's effectiveness. Poor etiquette can nullify any advantage the design of the mitigations afforded by providing cognitive state classification. Etiquette was addressed in both the IHMC and CMU evaluations and will need to be considered at every evaluation in the future. Another key issue is how the information is formatted and presented to the operator, which will determine the user acceptance and system success.

A poorly formatted display or clunky interactions will not be overlooked by the target population. Display design and interactions with the system need to be considered as the system is fielded with real operators.

System feedback is another important component of overall system effectiveness. System function transparency, or insight into what the system is doing and why, will directly affect the user's trust and acceptance of the system. Effort needs to be applied to appropriate levels of feedback from the system to the operator.

Knowledge of and agreement with the entity that has control of the system (the automation or the operator) has been the focus of human factors and human-computer interaction studies since the onset of automated systems. AugCog systems will be no different, and the operator's trust and interaction with the system will be directly related to his or her belief in whether the appropriate entity has control. Ultimately, the user will want to ensure that he or she is in control, and the interaction of the AugCog components should be so seamless that any changes in the system will be anticipated and welcomed by the user.

5 Augmented Cognition Program Phase 3

5.1 Phase 3 Introduction

5.1.1 Phase 3 Research Team

The Honeywell Augmented Cognition (AugCog) team in Phase 3 consisted of the collaborative efforts of Honeywell Laboratories, Advanced Brain Monitoring, Drexel University, and Oregon Health and Sciences University. In addition, the team was advised by the Natick Soldier Research, Development and Engineering Center (NSRDEC). Phase 3 of the program encompassed work done from January 1, 2005, though December 31, 2005.

5.1.2 Phase 3 Research Objectives

The Phase 3 Spring Cognitive Validation Experiment (CVE) was the next in a series of planned evaluations for aspects of the Honeywell team's closed-loop integrated prototype (CLIP) in an outdoor field environment. The Honeywell effort was concerned with mitigating high workload demands in the dismounted Soldier environment, especially with regard to information overload due to netted communications. This particular effort demonstrated a CLIP that integrated a kernel-based classification of cognitive state with an adaptive system designed to maintain high levels of performance under increasing workload. The effectiveness of the classification algorithms was evaluated to detect the user's cognitive state by correlating classification output to performance in various task load conditions. The team investigated the effectiveness of the mitigation strategies, the Communications Scheduler, and the Tactile Navigation Cueing System, to modify tasks based on cognitive state and thereby influence overall performance.

5.2 Phase 3 Challenges

The work in Phase 3 was motivated by several challenges (operational, algorithmic, and evaluative) that were addressed to move the technology from the lab to the field.

5.2.1 Operational Definition of Stress

Physical exertion is one of the primary stressors an Army Soldier faces on the battlefield. Simply moving to a rally point in a mission is made difficult when it requires the Soldier to carry an 80- to 120-lb. load (Girolamo, 2005). Other common stressors that can diminish cognition include heat (Steinman, 1987; Buller, et al. 2005), cold, limited food and water (Mountain, Sawka, & Wenger, 2001; Buller, et al., 2005), fear, and sleep deprivation. Stress will affect all aspects of information processing, including general arousal, selective attention, speed and accuracy performance, and working memory (Hockey, 1986). The degradation in cognitive performance that often results from the effects of stress can have catastrophic results, such as the poor decision-making in the Vincennes incident when the crew of the U.S. vessel mistakenly identified a civilian airliner as a hostile aircraft and shot it down (U.S. Navy, 1988; APA Monitor, 1988).

5.2.2 Classification

Inferring cognitive state from noninvasive neurophysiological sensors is a challenging task even in pristine laboratory environments. Artifacts ranging from eye blinks to muscle artifacts and electrical line noise can mask electrical signals associated with cognitive functions. These concerns were particularly pronounced in the context of the Honeywell team's efforts to realize neurophysiologically driven adaptive automation for the dismounted ambulatory Soldier. Besides the typical sources of signal contamination, the Honeywell team accounted for the artifacts induced by shock, rubbing cables, and gross muscle movement.

This chapter presents the Honeywell team's efforts to make reliable sensor-based cognitive state assessments given the constraints just cited. Described below is a system designed to facilitate cognitive state classification in mobile environments. The hardware configuration allowed neurophysiological data to be collected and processed in a body-worn wireless platform. An overview of software components used for signal processing and artifact reduction is described with an emphasis on the classification approach. Additionally, validation results indicate that it was feasible to discriminate among workload levels on the basis of neurophysiological sensors in ambulatory contexts.

Realizing the vision of the AugCog program in the context of an ambulatory Soldier was constrained by several challenges. First, as Schmorrow and Kruse (2002) have noted, processing and analysis of neurophysiological data is largely conducted offline by researchers and practitioners. However, for AugCog technologies to work in practical settings, effective and computationally efficient artifact reduction and signal processing solutions were necessary. Second, inferring the cognitive state of users demanded pattern recognition solutions that were robust to noise and the inherent nonstationarity in neurophysiological signals. Third, it required the development of means to collect reliable neurophysiological data outside the laboratory. Hence, compact and robust form factors associated with neurophysiological sensors and processors were a matter of critical concern. Users should be able to move around freely.

5.2.3 Mitigation

Design of scenarios to empirically assess both the classification and performance enhancement capabilities of the Honeywell AugCog system in a mobile environment was pursuant to a multitude of sometimes contrary constraints. Tasks must be:

- Classifiable
    - Tasks (or resulting cognitive state) must be reliably detected by a cognitive state assessor (CSA).
    - The researchers must understand how the task load affected the participant's workload.
- Augmentable
    - Tasks must enable the researchers to augment performance when the participant is under stress.
- Relevant

    - Tasks must be relevant to the Army and the Future Force Warrior (FFW).
    - Tasks must be consistent with the roles chosen (e.g., platoon leader).
- Feasible
    - Enough experiment control must exist to enable proper assessments.
    - Assessments must be repeatable within constraints of participant limitations.
- Measurable
    - A continuous data stream of performance metrics would be preferable.
    - Ways to adjust the workload between high and low by manipulating the task load would be desirable.

Given these constraints, the experiment had certain limitations and assumptions. For the mitigations, certain simplifying assumptions were made in the interest of rapid prototyping. For instance, the Communications Scheduler depended on the notion of message priority. To date, these priorities have been preset and fixed with each message. However, in a fielded system, the priorities would have to adhere to military doctrine as well as being modifiable by the human operator (see Section 5.3.2.3 for more discussion). The Tactile Navigation Cueing System implemented a very simple waypoint-to-waypoint navigation system. A fielded system would have to update the "ideal" path constantly to take into account current location, obstacles, and destination.

5.3 Phase 3 System Design and Architecture

Details of the Phase 3 CLIP configuration can be found in Appendix D. This section describes some of the principal components of the CLIP.

5.3.1 Cognitive State Assessor

5.3.1.1 Signal Processing Software

The cognitive state classification efforts reported here relied primarily on Electroencephalogram (EEG) data. As mentioned earlier, the sensor monitoring equipment consisted of a BioSemi Active Two-EEG system with 32 electrodes. Vertical and horizontal eye movements and blinks were recorded with electrodes below and lateral to the left eye. All channels referenced the right mastoid. EEG was sampled at 256 Hz from seven channels (CZ, P3, P4, PZ, O2, P04, F7) while the participant was performing tasks. These sites were selected based on a saliency analysis on EEG collected from various participants performing cognitive test battery tasks (Russell & Gustafson, 2001). EEG signals were preprocessed to remove eye blinks using an adaptive linear filter based on the Widrow-Hoff training rule (Widrow & Hoff, 1960). Information from the VEOGLB ocular reference channel was used as the noise reference source for the adaptive ocular filter. DC drifts were removed using high-pass filters (0.5-Hz cut-off). A bandpass filter (between 2 and 50 Hz) was also employed, as this interval was generally associated with cognitive activity. The overall schematic diagram of the signal processing system is shown in Figure 44.

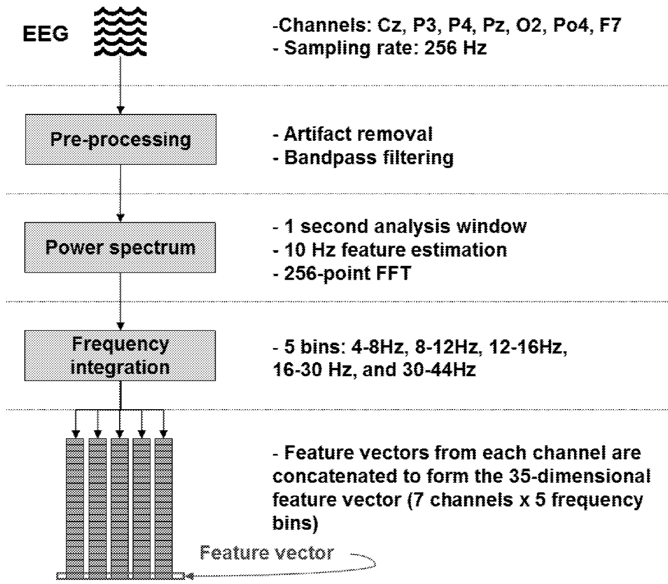


Figure 44. Signal processing system.

The power spectral density (PSD) of the EEG signals was estimated using the Welch method (Welch, 1967). The PSD process uses 1-second sliding windows with 50% overlap. PSD estimates were integrated over five frequency bands: 4-8 Hz (theta), 8-12 Hz (alpha), 12-16 Hz (low beta), 16-30 Hz (high beta), 30-44 Hz (gamma). These bands, sampled every 0.1 second, were used as the basic input features for cognitive classification. The particular selection of the frequency bands was based on well-established interpretations of EEG signals in prior cognitive and clinical (e.g., Gevins, Smith, McEvoy & Yu, 1997) contexts.

5.3.1.2 Cognitive State Classification System

Estimates of spectral power formed the input features to a pattern classification system. The classification system used parametric and nonparametric techniques to assess the likely cognitive state on the basis of spectral features, i.e., estimate p (cognitive state spectral features). The classification process relied on probability density estimates derived from a set of spectral samples. These spectral samples were gathered in conjunction with tasks representative of the eventual task environment. These sample patterns were assumed to be representative of the population of spectral patterns one would expect in the performance environment. The classification system used three distinct classification approaches: K-nearest neighbor (KNN), Parzen Windows, and Gaussian Mixture Models (Figure 45).

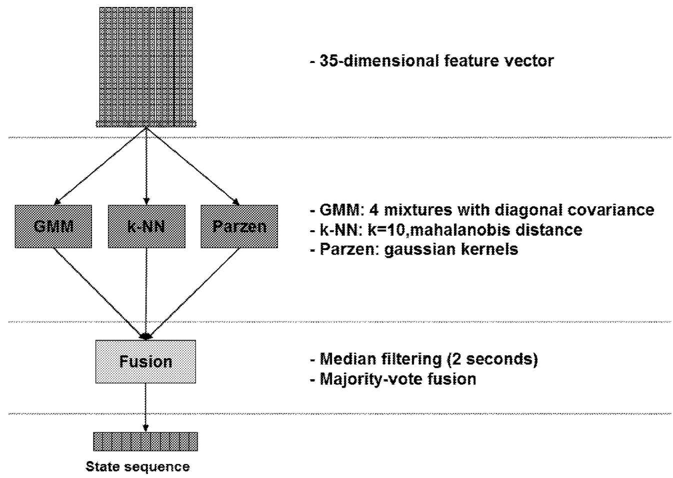


Figure 45. Classification system.

Gaussian Mixture Models: Gaussian mixture models (GMM) provided a way to model the probability density functions of spectral features associated with each cognitive state. This was accomplished using a superposition of Gaussian kernels. The unknown probability density associated with each class or cognitive state was approximated by a weighted linear combination of Gaussian density components. Given an appropriate number of Gaussian components, and appropriately chosen component parameters (mean and covariance matrix associated with each component), a Gaussian Mixture Model could model any probability density to an arbitrary degree of precision.

The parameters associated with component Gaussians were iteratively determined using the Expectation Maximization algorithm (Dempster, Laird, & Rubin, 1977). Once the Gaussian parameters had been initialized, the system iterated through a two-step procedure for each sample associated with each class. In the first step (expectation step), the system computed the probability of a particular training sample belonging to a particular class based on current model parameters (posteriori probability). In the maximization step, the model parameters were adjusted in the direction of increased class membership likelihood.

Once probability density functions associated with each cognitive state had been generated, it became possible to classify individual spectral samples. Each spectral vector was attributed to a class that has the highest posterior probability of representing it. Posterior probabilities were computed using Bayes' rule. Figure 46 shows the probability density functions associated with three distinct classes. These probability densities were estimated using three Gaussians. For example, very high values of the data point x were most likely to come from class 3, and very low values of x were most likely to come from class 1.

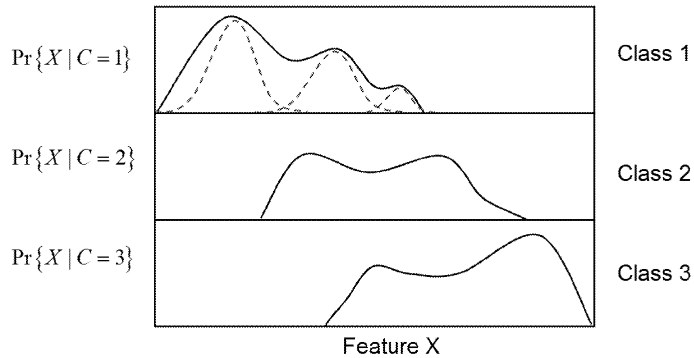


Figure 46. Gaussian mixture models.

K Nearest Neighbor: The K-nearest neighbor approach was a nonparametric technique that made no assumption about the form of the probability densities underlying a particular set of data. Given a particular sample x, the classification process identified k samples whose features come closest (as assessed by Euclidian or Mahalanobis distance metrics) to the features represented in x. The sample x was assigned the modal class of the nearest k neighbors. For example, consider the data point represented by the question mark in Figure 47. Based on k = 5, it would be assigned the label associated with the most common class category of its five nearest neighbors: 1. It can be shown that if k is large, but the overall cell small, that the classifier will approach the best possible classification (Bayes rate) (Duda, Hart, & Stork, 2000).

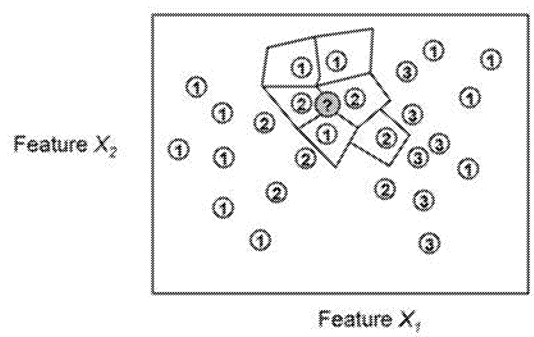


Figure 47. K-nearest neighbor.

Parzen Windows: Parzen windows (Parzen, 1967) were a generalization of the K-nearest neighbor technique. Instead of choosing the nearest neighbors and assigning a sample x with the label associated with the modal class of its neighbors, one could weight each vote by using a kernel function. With Gaussian kernels, the weight decreased exponentially with the square of the distance. As a consequence, far-away points became insignificant. Kernel volumes constrain the region within which neighbors were considered. Consequently, Parzen windows were a better choice when there were large differences in the variability associated with each class. The data point shown in Figure 48 was assigned to the dominant class in its immediate vicinity.

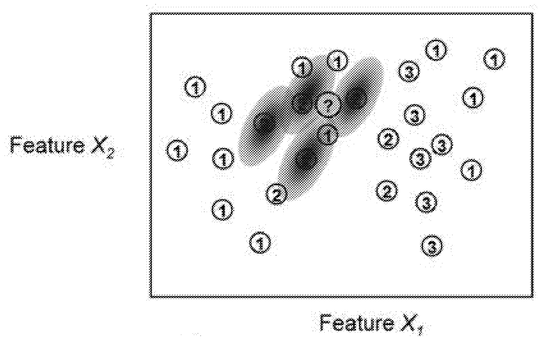


Figure 48. Parzen windows.

Composite Classifier: These statistical classification techniques were chosen over multi-layer neural networks because they required minimal training time. KNN and Parzen Windows required no training, whereas the Expectation Maximization (EM) algorithm used to generate GMMs converged relatively quickly. KNN and Parzen windows approaches required all training patterns to be held in memory. Every new feature vector had to be compared to each of these patterns. However, despite the computational cost of these comparisons at runtime, the system was able to output classification decisions well within real-time constraints.

The composite classification system regarded the output from each classifier as a vote for the likely cognitive state. The majority vote of the three component classifiers formed the output of the composite classifier. When there was no majority agreement, the Parzen windows decision was selected. A classification decision was output at a rate of 10 Hz. Outputs from the composite classifier were passed through a modal filter before an assessment of cognitive state was output by the classification system. Modal filtering served to make the cognitive state assessment process more robust to undesirable fluctuations in the underlying EEG signal. Modal filtering was done over a sliding 2-second window with the assumption that cognitive state remained stable over that period of time.

5.3.1.3 Validation Study

Validation experiments compared classification accuracy across three task load levels in two mobility conditions: stationary and walking. The tasks in the stationary case were: *relaxed* (waiting for orders), *communicate* (getting orders from base via radio communication), and *count* (starting from 100 and decreasing by 7). Tasks in the mobile case were *navigate* (walking to a designated target), *navigate and visual search* (walking while looking for snipers), and *navigate and communicate* (receiving and giving mission status reports). The participant wore the sensor suite described earlier in this report in both mobility conditions. EEG was collected as participants performed each of the tasks mentioned above.

After the preprocessing and PSD feature extraction stages, approximately 3,000 samples were obtained. One-third of this data was used for training the classifiers, and the remaining two-thirds were used for testing. Classification results for both stationary and mobile cases are presented in the confusion matrix shown in Figure 49; higher numbers on the diagonal of each matrix correspond to better performance. As the diagonals associated with each confusion matrix indicate, classification accuracy was well over

90%. The results presented here were representative of outcomes replicated with a large number of independent data sets and cognitive tasks.

STATIONARY

true class \ classified as	Relaxed	Communicate	Count
Relaxed	1.000	0.021	0.000
Communicate	0.000	0.979	0.098
Count	0.000	0.000	0.902

MOBILE

true class \ classified as	Navigate	Search	Nav & Comm
Navigate	0.959	0.019	0.000
Search	0.003	0.981	0.047
Nav & Comm	0.038	0.000	0.953

Figure 49. Probability of classifying test patterns correctly.

5.3.2 Mitigation Strategies

There were four broad categories of possible mitigations in an AugCog system:

- Task/Information Management
- Modality Management
- Task Offloading
- Task Sharing

Two principal mitigations were employed by the Augmentation Manager; each addressed a different task found in the dismounted Soldier domain. Table 30 describes how each mitigation strategy (The Communications Scheduler and the Tactile Navigation Cueing System) relates to the scenarios found in the Spring CVE.

Table 30. Classes of mitigation strategies.

Mitigation Strategy	**Scenario 1** Communications Scenario	**Scenario 2** Navigation Scenario
Task Scheduling	Communications Scheduler	
Task Offloading		
Task Sharing		Tactile Navigation Cueing
Modality Management	Communications Scheduler	Tactile Navigation Cueing

In Scenario 1, the primary mitigation was Task/Information Management via the Communications Scheduler. In addition, the Communications Scheduler's ability to change audio messages to text was a form of modality management as well.

In Scenario 2, the system utilized a tactile display via the Tactile Navigation Cueing System, to assist the participant in the navigation to the safe zone task.

5.3.2.1 Expectations

The Spring 2005 CVE focused on stressing the task components that involve attention for information processing, and likewise the mitigations were focused primarily on the attention bottleneck. In Scenario 1, the problem was the fact that netted communications produced more information than the Soldier could handle. This was mitigated by

scheduling messages by priority, task information, and cognitive state. In addition, the attention bottleneck in Scenario 1 was mitigated via modality management—changing from audio messages to lower priority, deferred text messages. The Communications Scheduler mitigation was expected to enhance performance on the primary, high-priority task by focusing attention on those high-priority communications while deferring messages related to lower priority tasks. Cognitive state assessment determined when the Soldier was overloaded, and required the Communications Scheduler mitigation to intervene in the message flow in an appropriate manner.

The sensory input bottleneck was encountered in Scenario 2, where the participant was overloaded in the visual and auditory domains. This was mitigated via modality management to utilize the underused tactile domain. In addition, the Tactile Navigation Cueing System was expected to change the processing demands associated with the "Navigation Through Unfamiliar Area" task from a primarily cognitive task (involving reading a map, mental transformation from 2-D to 3-D space, etc.) to a primarily reactive task (responding to tactile cues), thus freeing up cognitive resources for accompanying tasks.

5.3.2.2 Automation Etiquette

The pros and cons of automating complex systems have been widely discussed in the literature (e.g., Parasuraman & Miller, 2004; Sarter, Woods, & Billings, 1997). Automated systems have brought precision and consistency to tasks, relieved operator monotony and fatigue, and contributed to economic efficiency. However, as widely noted, poorly designed automation has had serious negative effects. Automation could relegate the operator to the status of a passive observer, serving to limit situational awareness, and induce cognitive overload when a user may be forced to inherit control from an automated system. Norman (1992) has suggested that most problems associated with complex automated systems stem from the poor feedback that many systems provide to users. He has argued that it is possible to reduce error through appropriate design considerations:

> *Appropriate design should assume the existence of error, it should continually provide feedback, it should continually interact with operators in an effective manner, and it should allow for the worst of situations. What is needed is a soft, compliant technology, not a rigid, formal one* (Norman, 1992).

Automation technologies that have emerged under the AugCog program address certain aspects of Norman's prescriptions for the design of automated systems. They offer the potential to engage the user in a mixed-initiative interaction—leveraging the strengths of both machines and their human operators. Based on real-time assessments of cognitive state, these systems dynamically provided assistance to users when they are likely to be overwhelmed by task demands. However, there are several features of neurophysiologically triggered automation that can have a detrimental effect on performance. First, many neurophysiological indices fluctuate rapidly over short time windows. Triggering automation on the basis of an index with a high degree of inherent nonstationarity can severely disrupt task performance. Second, adaptive assistance can alter the task demand that the controller is subject to. As a consequence, neurophysiological measures may not effectively reflect the overall task demand imposed by

the task environment. Unless the task context is assessed and considered using non-physiological sensors, a neurophysiologically triggered adaptive system could potentially return control to the user under circumstances that may be beyond the capability of a user to handle. Third, despite the fact that systems developed under the AugCog program display a high degree of sensitivity to a user's cognitive state, as automated systems, they stand to inherit many of the problems commonly observed with highly automated human-in-the-loop systems.

The following sections describe the two principal mitigation strategies employed in the Spring CVE: the Communications Scheduler and the Tactile Navigation Cueing System. In addition, they describe the considerations that went into the design of the Honeywell AugCog system to help dismounted Soldiers in a mobile environment perform effectively under extreme cognitive demands. Specific design decisions are described that address many of the concerns raised by researchers to adequately define both the cost and the benefits of the application of automation assistance in context.

5.3.2.3 Communications Scheduler

The Communications Scheduler mitigated the attention bottleneck via task scheduling and modality management of incoming communications. The system was tasked with determining when and how information was displayed to the Soldier. The Communications Scheduler scheduled and presented messages to the Soldier based on the cognitive state profile (CSP), the message characteristics, and the current context (tasks). Based on these inputs, the Communications Scheduler passed through messages immediately, deferred and scheduled nonrelevant or lower priority messages, escalated higher priority messages that were not attended to, diverted attention to incoming higher priority messages, changed the modality of message presentation, or deleted expired or obsolete messages.

Message Characteristics: All messages had a priority associated with them, depending on how critical they were. Current military radio operations do not have a priority embedded within the message; however, some digital communications technologies have been proposed, including a priority associated with the messages based on the FIPR (Flash, Immediate, Priority, Routine) scheme. For instance, the Army currently fields an information system, FBCB2 (Force XX1 Battle Command Brigade and below), that is used across echelons from vehicle commanders up through the battle command staff. The FBCB2 system uses the FIPR prioritization scheme for incoming messages (Durlach, 2004).

When the mitigation was in effect, messages were scheduled according to certain rules. High-priority messages were mission-critical and time-critical, which means they must have been heard and understood as soon as they arrived. Medium-priority messages were mission-critical, but had a larger time window to work with. A medium-priority message was potentially deferred if the system found that the participant was highly engaged in another, higher priority task. All medium-priority messages were played before the end of the mission. Low-priority messages were not mission-critical or time-critical. They were presented if the participant was not engaged in another task. If the system found that the participant was engaged in another task, the low-priority message was presented in text format in the message window of the Soldier's personal digital assistant (PDA).

Message Alert Modes: High-priority messages had a tone played once before they were presented. If the system found that the participant was highly engaged in a task, it would play the louder and more salient tone once before the message was presented. Medium-priority messages also had a tone played once before they were presented. This tone was recognizably different than the high-priority tone. Low-priority messages did not have a tone associated with them.

System Logic: The CSA determines the CSP decision variable: *Workload*. The Communications Scheduler determines the initial message presentation based on a user's current *Workload*. The Communications Scheduler performed one of three actions when it decided how to first present the message:

- Presented the message immediately in the audio modality with the appropriate "normal" tone preceding it.
- Presented the message immediately in the audio modality preceded by the appropriate "higher saliency" tone.
- Presented the message immediately in the text modality on the participant's Tablet PC.

The decision logic of the Communications Scheduler is summarized in Table 31. Each Workload cell had a rule P(modality, saliency), where P = play, modality = audio or text, and saliency = normal, higher.

Table 31. Communications Scheduler decision rule set, where each rule is of the form play (modality, saliency).

	Before first message presentation		
CSP Variable Priority	High	Medium	Low
Workload High	P(audio,higher)	P(text,normal)	P(text,normal)
Workload Low	P(audio,normal)	P(audio,normal)	P(default,normal)
Workload Low After High	P(audio,higher)	P(text,normal)	P(text,normal)
Workload Unknown	P(audio,normal)	P(audio,normal)	P(audio,normal)

The Message Application: When invoked, the Communications Scheduler deferred low-priority messages to a text display on the PDA (see Figure 50). Messages were ordered by priority first and time second. Thus, all high priority messages appeared at the top, and within a single priority, the most recent message appeared at the top. All messages appeared, regardless of whether they were also presented over the radio. This was to avoid any confusion an incomplete or sporadic recording of messages may induce. The participant could click on the message in the list, and the full message appeared below in the Message Details Box. Unread messages were in boldface in the message list. Messages were in boldface until they were read. Clicking an individual message would "unbold" it and would indicate that it had been read. When finished the participant would indicate all messages had been read by pressing the "Read All" button.

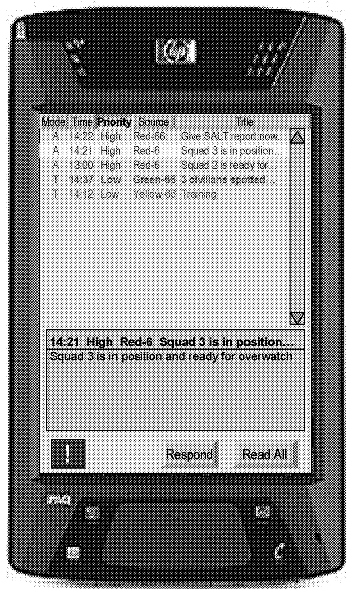


Figure 50. The Message Application on the PDA.

Automation Etiquette: Research shows that unless users were able to predict clearly how an automated system was likely to perform, automation may introduce more problems than it solves (Sarter, Woods, & Billings, 1997). The mitigation strategies described here had very clear rules for eliminating uncertainty and unpredictability. The Communications Scheduler benefited users by allowing them to defer responses to messages under conditions when attention had to be split between competing tasks, thus allowing them to focus on higher priority tasks first. However, this kind of automated system behavior has negative side effects: Loss of momentary situation awareness and lags in responses could break coordination among teams and introduce inefficiencies in the mission. Thus, it was important that the Communications Scheduler be invoked only when the benefits of its use outweighed its costs. For that reason, the Communications Scheduler was not be used continuously, but only in times of high cognitive stress on the user, when faced with competing tasks that overload his or her ability to comprehend and process all incoming information.

Since the Communications Scheduler was not be used continuously, the issue of automation etiquette became important. The Communications Scheduler should be invoked (and should cease) in a manner that does not exacerbate confusion. The Communications Scheduler mitigation was invoked when workload was high—for instance, low-priority messages were deferred to the PDA. However, when workload dipped below the threshold used to trigger the message deferral, the Communications Scheduler continued to defer messages. This was because deferring communications on the basis of moment-to-moment fluctuations in gauge values could be confusing. Messages could be misinterpreted without surrounding context if they were to be played in audio modality after their predecessor messages were deferred to the PDA (and remain unread for a period of time). If expected messages were not heard, it may have been hard to disambiguate whether this was because of the Communications Scheduler or some mission-related cause. To avoid confusion, once communications scheduling was activated, all low- and medium-priority messages were deferred to the PDA until the user

caught up on all messages and clicked a "messages read" button (Mathan, Dorneich, & Whitlow, 2005).

5.3.2.4 Tactile Navigation Cueing System

While in hostile territory, Soldiers have to be able to adapt their navigation plans to evolving tactical threats. However, during engagements in hostile territory, the cognitive resources necessary to generate a safe route while engaging the enemy and handling communications may simply not be available. To address these concerns, the Honeywell AugCog prototype incorporated functionality to assist users with navigation tasks. In hostile areas, the system generated navigation plans based on knowledge of the mission's geographical objective and information about enemy locations gathered from the FFW communications networks. The system provided users with a graphical plan in conjunction with tactile cues to guide them through relatively safe zones. Navigation assistance was invoked when the CSP indicated workload was high and the participant needed to navigate through an unfamiliar route.

System Description: Tactile cues were provided to the user by means of a tactor belt worn around the waist. Tactors were fired to direct participants toward the bearing of their next waypoint The rate of firing the tactors increased from 1 to 2 to 8 Hz as the participant approached each waypoint. When a waypoint was reached, the system provided navigation cues relative to the next waypoint until the participant reached the appropriate destination.

System Logic: In the unmitigated version of Scenario 2, the participants were required to refer to their map, orient themselves to the current location, and determine the next-best route to the safe zone and avoid being ambushed. In the mitigated scenario, the participants received tactile cues that guided them in the correct direction to reach the safe zone. Thus, the navigation task went from being cognitively intense to essentially a reactionary task to external stimuli. This was designed to lower the task load and cognitive demands, allowing participants to improve performance on the navigation task while not adversely affecting other tasks being done simultaneously. Tactile cues have been shown to be effective in improving performance of spatial tasks, even in the presence of competing secondary workload tasks (Raj, Kass, & Perry, 2000).

Automation Etiquette: Operationally, pulses from the tactor belt "tugged" the participants in direction they were expected to go. The system was invoked when the CSP indicated workload was high and the participant needed to navigate through an unfamiliar route. However, turning the system off as soon as workload fell below some threshold would have left users disoriented in an unfamiliar area. Thus, once the system was turned on, the navigation mitigation persisted until users arrived at the safe destination. The system was only used when workload was high, rather than any time the participant needed to navigate through an unfamiliar route, since there was a potential loss of situation awareness when the participant was not forced to navigate on his/her own. This cost to the mitigation may never have been realized. However if the Soldiers were to find themselves in the area at a later time or their commanders assumed they knew their way around because they had been there before, the lack of knowledge of the area could have had detrimental effects. Although this loss of situation awareness was acceptable in high

task load situations, it was an unnecessary cost when Soldiers were capable of navigating the area on their own.

5.3.2.5 Cost/Benefit Approach

Although the mitigation strategies just described promised to help users perform critical tasks under extreme task contexts, as with any complex automated system, they had the potential to hurt task performance in a variety of ways. The system described here was designed with close consideration of several unexpected problems with human-automation interaction, as highlighted by Sarter, Woods, and Billings (1997). This section describes each of these potential problems and summarizes the design features they motivated.

Uneven Distribution of Workload: As Sarter and colleagues (1997) pointed out, many automated systems actually hinder performance in high-workload conditions. Many systems required the user to play the role of a translator or mediator—communicating aspects of the task environment to the system. Operators have had to take on the responsibility of explicitly specifying task parameters for the automation to execute. In many cases, these demands came during the busiest phases of work.

Automation in the context described here was designed to be invoked and parameterized with minimal involvement from the user. The mitigation strategies described here were triggered based on assessments of cognitive state and task context. As a result, users received automated assistance automatically in difficult task contexts—users were not distracted from the task at hand to configure the automation's intervention. Parameterization of the automated system was supported by the assumed netted communications infrastructure that was a central component of the FFW program. For example, likely ambush locations were assumed to be continually assessed using information from human and electronic surveillance assets. Real-time access to this information by the mitigation engine would allow the system to come up with route plans without explicit intervention from the user.

Breakdowns in Mode Awareness: Sarter et al. (1997) defined mode awareness as the ability of a system user to anticipate the behavior of automated systems. They suggested that breakdowns in mode awareness, so-called automation surprises, could lead to errors of omission in which the operator failed to observe and respond to uncommented or undesirable system behavior.

There were several sources of potential automation surprises in the context described here. First, neurophysiological and physiological indices that served to invoke automation embody a great deal of inherent nonstationarity. Triggering mitigations on the basis of signals that vary a great deal over short time windows could have been extremely disruptive for the user. To address this problem, cognitive state classification was based on joint consideration of several indices. Some of the indices employed in Honeywell's current and past systems included EEG, galvanic skin response, heartbeat variability, and pupilometry. These redundant sources of information combined to provide a relatively more stable indication of cognitive load than any single index would. Efforts during the spring CVE added additional robustness to cognitive state classification by choosing the modal classification output over specified time windows. The tradeoff between

mitigation latency and required classification robustness determined the size of window employed.

Second, once effective mitigation strategies were triggered, they effectively reduced the cognitive load on the user. Consequently, neurophysiological and physiological indices lost their value as indicators of task load. Disengaging mitigations solely on the basis of indices associated with cognitive load could return control to users under very difficult task conditions. To address this issue, mitigations were turned off on the basis of context-related information (i.e., information that was independent of cognitive state assessment). For example, communications scheduling was turned off after the users indicated that they caught up on all the deferred messages. Navigation cues were terminated only after a user had arrived at the destination, since stopping navigation cueing once it starts is likely to result in even greater disorientation due to an inherent loss of situation awareness (SA) when being cued versus navigating with a map. The loss of SA was deemed an acceptable cost when cognitive overload threatened a complete breakdown in performance. However, this cost must be accounted for when deciding how to "turn off" the mitigation. In this case, it was appropriate to continue migration until the destination was reached.

Third, the system described here provided a range of different types of assistance to users in different task contexts. Each of these mitigations assumed control over a certain aspects of a user's task. Unless users were clearly aware of the status of the adaptive system, they could have encountered a range of automation surprises. To avoid these problems, the system was always explicit concerning what mode the system was in. Once navigation aiding was turned on, users felt pulses on a belt that unambiguously conveyed the navigation mode to them.

New Coordination Demands: Sarter and colleagues (1997) suggested that autonomous automation components effectively became like crew members by taking over aspects of critical tasks. However, unlike good crew members, poorly designed automation may fail to keep users informed about task status. These systems may perform tasks autonomously and silently, but return control to users abruptly when things fail. This increased the coordination requirements and could have added to cognitive load.

Elements of the system described here were designed with the assumption that they were fallible. Mitigations should be designed to allow a human to intervene if the system is unable to handle a situation effectively. In the simplest example, users should have the ability to turn off the mitigation if it is not performing up to expectations.

Complacency and Trust in Automation: Sarter and colleagues (1997) suggested that complacency induced by automation may have been a critical factor in many accidents. They posited that users may have come to rely on automation, not realizing that these systems, although largely reliable, may have been fallible.

Issues of complacency were also of concern in the context of the effort described here. By delegating critical tasks such as communications and visual monitoring to an automated system, users faced the risk of missing critical task-relevant information. The approach to reducing possible complacency relied on training to emphasize the fallibility of the system and to provide users with procedures for monitoring the system and resuming control of delegated tasks as soon as practical.

Training: Automation of complex tasks often introduced the need for additional training. Besides learning to master the performance of inherently complex tasks, users had to learn about the use of complex automation components to support the execution of these tasks. Sarter and colleagues (1997) noted that sophisticated automation components interact with the task environment in complex ways. They argue that training must occur in the context of use for users to be able to acquire accurate mental models of the system.

All participants who used the system received extensive training in the use of automation components in the actual contexts of use. Participants progressed to task scenarios only after they were able to successfully demonstrate use of each automation component in training scenarios. Participants also had to answer a broad range of questions about each automation component and its interaction with the task environment.

Cost/Benefit Tradeoffs: Although the mitigations described here had the potential for boosting performance when human cognitive resources may be limited, they could have had detrimental effects if left on at all times. The benefits and costs associated with these mitigations are shown in Table 32. Gauge-driven mitigation allowed these mitigations to be activated when the benefits were likely to outweigh the costs.

Table 32. Costs and benefits of mitigations.

Mitigation Agent	Benefits	Cost
Communications Scheduler	Allows users to defer responses to messages under conditions when attention has to be split among competing tasks	Loss of momentary situational awareness Lags in responses could break coordination among teams and introduce inefficiencies in the mission
Tactile Navigation Cueing System	Automated navigation assistance to enable users to focus on other critical tasks that demand attention	Loss of situational awareness since user is passive in the navigation task. Cause of many accidents—such as the American Airlines crash in Cali, Columbia.

5.4 Phase 3 Concept Validation Experiment

5.4.1 Experiment Objectives

Augmenting the dismounted Soldier with direct measures of his or her cognitive state was expected to enhance overall performance by triggering mitigations such as priority-based task management and modality-appropriate information presentation. The fully equipped Soldier/participant was outfitted with a mobile, sensor-based ensemble that monitored his/her cognitive and attentional state and automation to adapt the human-automation interface. Of particular importance was the participant's ability to handle the continuous inflow of netted communications and to direct his or her attention to the highest priority task to complete the mission in a highly dynamic environment. This research aimed to address the following questions:

- Would the integrated sensor-driven classification of cognitive state detect a change in the Soldier's cognitive state between low task load and high task load conditions?

- Would cognitive state changes correlate with changes in performance?
- Would the Communications Scheduler mitigation strategy effectively alter the Soldier's cognitive/attentional state in order to focus attention and improve comprehension on the highest priority items?
- Would there be any cost to the use of the Communications Scheduler, such as a loss of situation awareness of lower priority message content?
- Would the tactile navigation cueing be intuitive to learn and successfully guide the participants to avoid danger zones and successfully reach the target area?
- Would there be a cost to the use of tactile cueing, such as a loss of situation awareness regarding information normally acquired en route?
- Would mitigated performance be superior to unmitigated performance?

5.4.2 Operational Scenario

There were two scenario types: a radio communications scenario and a movement to objective navigation scenario. Each participant served the role of a platoon leader and completed a total of four experiment trials, each with periods of low and high task loads.

The communication scenario required participants to navigate via a known (circular) and secure route while monitoring an ongoing mission, maintaining radio counts, and performing a periodic mathematical task.

The navigation scenario involved navigating a complex route to avoid video surveillance detection and virtual minefields while executing a mission to navigate to an objective to set up a fortified surveillance watch. Secondary tasks included scanning the environment for potential improvised explosive devices (IEDs), monitoring radio communications to organize another evolving mission, maintaining radio counts, and performing a periodic mathematical task that simulated effortful interruptions faced by dismounted leaders.

Each scenario was performed twice: once under a mitigated condition and once under an unmitigated condition. The order of presentation of mitigated and unmitigated trials was counterbalanced. Each scenario contained periods of high task load and periods of low task load. The scenario sets (two communications scenarios and two navigation scenarios) were similar, but not identical, to avoid any learning effects. The experiment session began with the participant being briefed on the current mission. The briefing presented all the information the participant needed to execute the mission, including descriptions of mission objective, overall tasks, a general description of the performance goals, and the time constraints for completing the mission. Upon completion of the briefing, the participant executed the first mission.

The participant performed a variety of tasks in the two scenarios (see Table 33).

Table 33. Tasks performed by participant in each scenario.

Task	Communications Scenario	Navigation Scenario
Navigate	Simple, known route	Complex, unknown route
Maintain Counts	X	X
Mission Monitoring	X	X
Tertiary Math Task	X	X
Visual Scan for IEDs		X
Maintain Situation Awareness	X	X

Unique mitigations supported task performance in each scenario once high workload was detected by the neurophysiological system. For the navigation scenario, the system provided navigation aiding via vibrotactile directional feedback. In the communications scenario, the system enabled message scheduling that deferred low-priority messages to a lower workload period.

Scenarios were run in a large, grassy field surrounded by light forest situated behind Honeywell Labs in Northeast Minneapolis, Minnesota, as seen in Figure 51. Participants interacted primarily with a handheld radio and a PDA. Input for the mission monitoring and counts tasks came over the radio, and the participants responded over the radio as well. A math interruption task was completed on a PDA. Within each scenario blocks of high and low task load conditions lasted approximately 5 minutes and 2 minutes, respectively. The primary difference between high and low task load periods was the pace of radio communications. The math interruption task occurred with equal frequency under both task load conditions.

Figure 51. Mobile system during testing.

5.4.2.1 Description of Navigation Scenario

Navigation Task (mitigated task in this scenario): in unmitigated trials participants navigated a complex route, with the benefit of a paper map, while avoiding detection by video surveillance cameras and areas with mines, both of which were represented on their map. In addition, participants scanned their environment to detect IEDs, which were not indicated on the map, and reported their location over the radio. Based on visible cues

and integration of map-based awareness, participants determined the safest route to travel. Coming in contact with any of the "forbidden" zones was detected via a dead reckoning module / global positioning service (DRM/GPS) device and resulted in a time penalty. The participants were alerted of incursions via aural alert tones. In mitigated trials participants wore eight-tactor vibrotactile devices around their waists, known as a Tactabelt (manufactured by Anthrotronix, Inc). Under the mitigated condition, the Tactabelt system provided vibrotactile navigation support by "buzzing" one of the eight tactors in the belt corresponding to the direction of the next waypoint along a predetermined safe path through the hazards. Once participants arrived at a given waypoint, the system buzzed all the tactors to indicate arrival at the current waypoint and that the next navigation buzz would be directed toward the next waypoint.

The complexity of this task resulted from the requirement that the participants divide visual attention between two visual information sources, the outside environment for IEDs and navigation and the near environment composed of the map and PDA device. The intent of adding the vibrotactile cueing device was to reduce the need to visually attend to the written map and process the navigation task via the tactile modality. It was hypothesized that this mitigation would improve performance on the navigation task and improve performance on other secondary tasks due to the availability of more attentional resources.

Maintain Radio Counts Secondary Task: A simulated company commander relayed messages about entities encountered by his or her three platoon leaders (PLs) over the radio. The participant was one of those PLs. The messages contained reports of civilians, enemies, or friendlies spotted. The participant maintained a running total of civilians, enemies, and friendlies reported to him or her while ignoring the counts reported to the other two platoon leaders. The task load was varied by the rate of incoming messages: six messages per minute under the high task load condition; two messages per minute under the low task load condition. During the 5-minute high task load period, participants were asked to report their counts five times; whereas, they reported twice in the 2-minute low task load period.

This task relied heavily on the participant's ability to keep the three counts in working memory until asked to report the counts. This task also required the participants to focus their aural attention to listen for their call-sign and ignore the messages directed to the other PLs.

Mission Monitoring Secondary Task: Each participant organized the execution of a series of bounded overwatch maneuvers by three squads under his or her command. In bounded overwatch, one squad moved while the other two protected the moving squad. Participants kept track of the status of all three squads —either "ready to move" or "ready for overwatch." Once all three squads reported that they were in position (two squads ready for overwatch and one squad ready to move), participants ordered the appropriate squad to move forward. In this task, the participant was responsible for keeping track of squad status and ordering the correct squad to move at the correct time (i.e., when all three squads reported that they were ready and in position).

This task required the participants to keep track of the three squads, their locations, and their readiness to advance in the mission in working memory until the final team was in

position. The nature of this task suggested there may have been a benefit to visualizing the movement of the virtual squad in time and space; therefore, the AugCog team speculated that this task was taxing spatial working memory even though the task was presented verbally.

Math Interruption Secondary Task: A single math problem was periodically presented to the participants as an interruption task during the scenarios. At the start of each interruption, a loud aural alert sounded on the PDA. The participant was required to acknowledge the alert with a click of the PDA stylus; at this point, a difficult math problem (adding two three-digit or two-digit numbers together) was presented, along with a 10-second countdown to put time pressure on participants. Participants entered the answer using a series of drop-down boxes. This task was representative of any type of unanticipated interruption that requires significant cognitive resources and an immediate response from the PL. Participants were interrupted twice per minute in both high and low task load periods.

This interruption task had the potential for disrupting any of the tasks that required continual rehearsal, such as the working memory tasks of mission monitoring and maintaining counts. Also, due to the head-down time with the PDA, it also had the potential for disrupting the visual search for the IEDs in the environment.

Visual Search for IEDs: The field in which the participant was navigating contained multiple IEDs. The IEDs were discs of various colors. Participants were instructed to radio in to report the sighting and approximate location of IEDs. This task forced participants to visually scan their environment.

Maintain Situation Awareness. Participants were required to maintain an awareness of their current location, the status of all teams and personnel reporting to them, the overall situation as relayed through radio communications, and their surroundings. Participants were asked to re-create the route they just took to move through the mine/camera field.

5.4.2.2 Description of the Communications Scenario

Navigation Task: The participants navigated along a familiar and marked route. The simple navigation in this scenario was required to increase task complexity, frame the mission in a multitasking environment, and test the performance of the neurophysiological and physiological sensors and cognitive state classifiers in a mobile environment.

Maintain Radio Counts Secondary Task (mitigated task in this scenario): This was the same task as in the navigation scenario; however, in high task load conditions, the Communications Scheduler was available. In the mitigated condition, the radio counts communications were deferred to the PDA to allow participants to total the counts once they completed the high task load tasks. This mitigation reduced the frequency of radio communications during higher task load periods and allowed participants to complete the counts under lower task load conditions.

Mission Monitoring Secondary Task: Same task as that used in the navigation scenario.

Math Interruption Secondary Task: Same task as that used in the navigation scenario.

Maintain Situation Awareness: As with the navigation scenario, participants were required to maintain an awareness of their current location, the status of all teams and personnel reporting to them, the overall situation as relayed through radio communications, and their surroundings. Participants were also asked about the content of low-priority messages they received.

5.4.3 Experiment Hypothesis

When the participants' tasks were augmented with a mitigation strategy for message scheduling and navigation, their performance on relevant tasks were expected to be enhanced without degrading their performance on nonrelated tasks. Performance enhancements included more accurate counts, more accurate target identification, faster response times to target identifications, and more accurate responses during mission monitoring due to the lower perceived levels of workload. Increases in task load (higher rates of information intake, more tasks requiring simultaneous attention) were expected to reduce performance. The mitigation strategy may only be effective in the higher task load conditions. The mitigations may also impose a cost to situation awareness for the task being mitigated. The navigational cueing may not require the participant to thoroughly process all the information in his/her surroundings due to the task assistance provided by the mitigation. There were three general hypotheses, one for each mitigation-based scenario and one for the classification approach.

Communications Scenario: The Spring CVE assessed the performance and workload effects for completing the primary task of mission monitoring and the secondary tasks of navigating, mission monitoring, maintaining counts, and responding to math problems. The experiment evaluated the effectiveness of the Communications Scheduler on the participants' overall performance on these four tasks. The CVE had the following hypothesis for the communications scenario:

> Hypothesis: *Scheduling of information would enhance the Soldier's performance on the counting task and mission monitoring tasks in high task load conditions without degrading performance on the remaining tasks.*

Navigation Scenario: The Spring CVE assessed the performance and workload effects for completing the primary task of navigating to the objective and the secondary tasks of mission monitoring and maintaining counts. The experiment evaluated the effectiveness of the tactile navigation cueing device on the participants' overall performance on these three tasks. The CVE had the following hypothesis for the navigation scenario:

> *Hypothesis: The use of tactile cueing during high-workload periods would enhance the participants' performance on the navigation tasks in high task load conditions.*

Cognitive State Classification Effectiveness: In addition to mitigation-related hypotheses, the Spring CVE assessed the effectiveness of the cognitive state classification approaches and the impact of mobility on classification performance. The classification algorithms used in the evaluation required the participant to be mobile in all scenarios. The sensors and output of the artifact removal algorithms are required to provide the classifiers with

data to discriminate between the low and high workloads during completion of the scenarios.

> <u>Hypothesis:</u> *There would be greater than 70% correct correlations between the neural net cognitive state classification output and the known levels of task load based on moment-to-moment classification.*

5.4.4 Experiment Design

Independent/Test Variables: Mitigation strategy (on/off) and Task Load (high/low). Each participant completed the two scenarios which contained periods of high and low task load in both the mitigated and unmitigated conditions.

Experiment Design: The communications scenario was a 2 (mitigation: mitigated, unmitigated) x 2 (task load block: high/low) within-participants design. The navigation scenario was a 2 (mitigation: mitigated, unmitigated) x 2 (task load block: low, high) within-participants design.

Task load Presentation: Each communications scenario had four task load blocks in a fixed order: high, low, high, low. The navigation scenario had two task load blocks in a fixed order: high, low.

Mitigation Counterbalancing: The presentation order of the mitigation was counterbalanced. The participant always received the communications scenarios followed by the navigation scenarios. See Table 34.

Table 34. Presentation order of mitigation in experiment trials.

Participant	P1	P2	P3	P4	P5	P6	P7	P8
Comm. Scenario 1	U	U	M	M	U	M	U	M
Comm. Scenario 2	M	M	U	U	M	U	M	U
Nav. Scenario 1	U	U	M	M	U	M	U	M
Nav. Scenario 2	M	M	U	U	M	U	M	U

Classification assessment was conducted by comparing the cognitive state classification accuracy across the low and high task load periods within each unmitigated block.

Performance enhancements produced by the mitigation strategies were assessed by comparing the unmitigated trials to the mitigated trials.

5.4.5 Dependent Measures

There were several objective and subjective dependent measures, as follows:

- Navigation: time to complete route: total number of incursions into zones with a security camera or land mine, successful completion of route reaching all the targets, composite accuracy/time metric (raw run time with a time penalty for each incursion).
- Maintain counts: reported vs. actual counts of civilians/enemies/friendlies, Accuracy metric.

- Mission monitoring: errors in squad to send forward, errors in timing of move command, composite accuracy metric (accounting for both squad errors and timing errors).
- Mathematical PDA task: response time to initiation alert, time to solve problem, accuracy metric.
- Visual scan for IEDs: percentage of IEDs reported.
- Maintain SA: percentage of SA questions answered (communications scenario), deviation between drawn path and actual path (navigation scenario).

Classification effectiveness:

- Correlations of task load manipulations within the scenarios with the output of the classification algorithm.

Workload measures:

- NASA-TLX (Task Load Index) rating scale: measures demands on a seven-index scale (mental, physical, temporal, performance, effort, and frustration). NASA-TLX was taken at the end of each experiment (task load) block.
- Post-experiment questionnaire.

Stress measures:

- Anxiety Rating Scale (ARS): measures cognitive state anxiety, somatic state anxiety, and self-confidence on a seven-point Likert scale.
    - ARS Question 1 was concerned with the stress induced by *performance anxiety*: I feel concerned about performing poorly, choking under pressure, and that others will be disappointed with my performance.
    - ARS Question 2 was concerned with the stress induced by *physical anxiety*: I feel jittery, my body feels tense, and my heart is racing.
    - ARS Question 3 ranked participants' *confidence* in their ability to perform well, where a higher rating equates to an increased confidence in performance: I feel comfortable, secure, and confident about performing well.
- Scale ranged from 1 to 7 (1 = not at all, 2 = a little bit, 3 = somewhat, 4 = moderately so, 5 = quite a bit, 6 = very much so, 7 = intensely so).
- ARS measures were taken at the beginning and end of each experiment block.

Subjective measures:

- SA Measures: post-experiment (high task load) block questionnaire to assess knowledge of low-priority messages (communications scenario). Knowledge of the route traveled was assessed after both blocks in the navigation scenario.

5.4.6 Participants

Eight participants completed the evaluation. All were male, between the ages of 20 and 42 (average age = 29.2), with 20/20 vision (normal or corrected), and normal audition.

Subjects were drawn from researchers and staff at the Honeywell Laboratory facility in Minneapolis, Minnesota.

5.4.7 Experiment Protocol

Training Trials: Two components of the training were conducted before the participants perform the experiment trials. The first training session was to ensure that all participants had a basic familiarity and proficiency with all the tasks they would perform in the experiment. To maximize the experimenters' time and the time spent collecting data on the day of the experiment, this training and all the paperwork associated with the evaluation were completed prior to the day of data collection. The participants also had a chance to practice the tasks on the day of the experiment. The second training session afforded the opportunity to collect data with which to train the cognitive state classifiers. Separate cognitive state classifiers were trained on the characteristics of the high and low task load periods, similar to those tasks found in the evaluation.

Experiment Trials: Four trials were performed. The order of the unmitigated and mitigated trials was counterbalanced, as shown in Table 34 (see Figure 52).

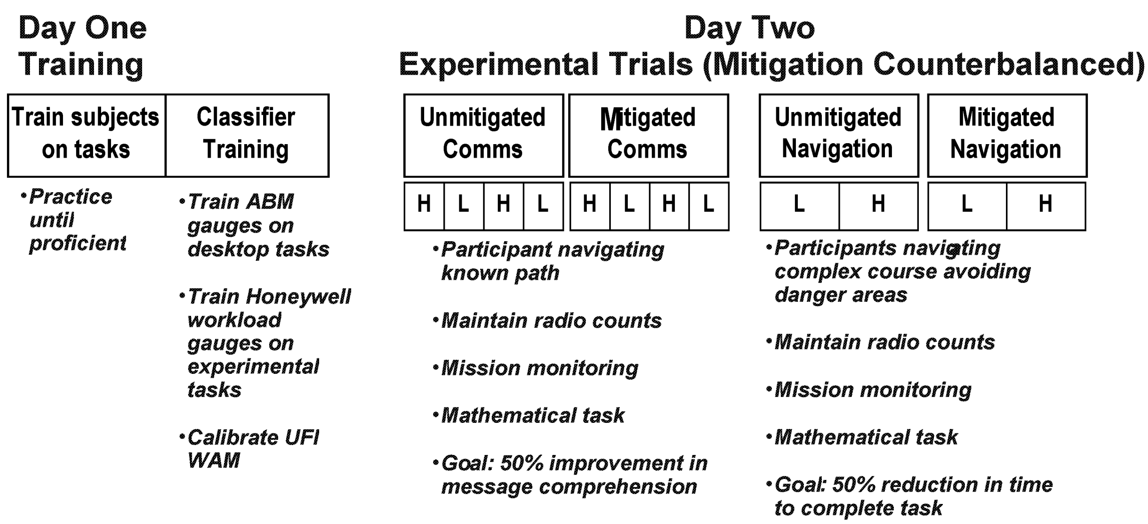


Figure 52. Experiment schedule.

Debrief: Following the completion of the experiment trials, participants were debriefed to obtain qualitative data regarding their experience, ability to complete the tasks, and effectiveness of the mitigation strategies.

5.4.8 Data Analysis Methodology

Two principal analyses were conducted in the Spring CVE: assessing the effectiveness of the mitigations and assessing the effectiveness of the classification analysis. In addition, the success workload manipulation designed in the experiment was assessed. Experimental conditions (independent variables of workload and mitigation) included: low-unmitigated, high-unmitigated, low-mitigated, and high-mitigated. Different combinations or subsets of these four conditions were used in particular analyses. Participants' performance was compared in the following conditions:

Communications scenario:

- *Mitigation effectiveness:* unmitigated vs. mitigated communications trials
- *Task load manipulation:* low task load periods vs. high task load periods in unmitigated communications trials
- *Mitigation benefit/cost:* unmitigated vs. mitigated trials for both high and low task load conditions in the communications trials

Navigation scenario

- *Mitigation effectiveness:* unmitigated vs. mitigated navigation trials
- *Task load manipulation:* low task load periods vs. high task load periods in unmitigated navigation trials
- *Mitigation benefit/cost:* unmitigated vs. mitigated trials for both high and low task load conditions in the navigation trials

Classification assessment was conducted by comparing cognitive state classification accuracy across the low and high task load levels within each unmitigated block.

5.5 Phase 3 CVE Results

Note that all graphs in this section report means and standard error. An alpha level of 0.05 was used on all statistical tests.

5.5.1 Cognitive State Classification Results

As part of the experiment, data collection was conducted with a six-channel EEG sensor headset made by Advanced Brain Monitoring, Inc. (ABM) and the 32-channel BioSemi system. The experiment setup supported real-time cognitive state classification by including training periods that emulated subsequent low- and high-workload conditions. After collecting between 5 and 10 minutes of EEG spectra data for both low- and high-workload training conditions, the data were submitted to the composite classification system to identify patterns to distinguish the workload conditions.

A crucial component of classification in field settings was a systematic procedure for selecting a subset of EEG features that was robust to potential artifacts and provides a basis for discriminating between workload classes. One way to do this was through an exhaustive selection of every possible feature combination drawn from the training data. The feature subset producing the best classification performance could have been selected for classifying cognitive state in the field. However, the cost of an exhaustive search was on the order of $O(2^n)$, where n represents the number of features. Thus, backward elimination (Langley, 1994), a heuristic search procedure through the space of possible feature subsets, was used to identify a subset of features that would provide reliable classification. Feature selection was based on the training data. With an appropriate selection of channels, the team was able to classify cognitive state with an accuracy that exceeded 70% for all participants. Accuracy was 95% for one participant. Performance with both the BioSemi (two participants) and ABM (six participants) systems was close to identical in the field environment. This finding was in contrast to the lab assessments, where the 32-channel BioSemi system provided better performance

relative to the six-channel ABM system. A possible explanation for this discrepancy may have been associated with differences in hardware design. The large number of relatively unconstrained cables associated with the BioSemi system may have been susceptible to movement-induced vibration, which may have been a potential source of noise. Any benefits of the additional channels the BioSemi system provides may have been lost to vulnerabilities to movement artifacts. In contrast, the ABM system was specifically designed for mobile use. If these results can be replicated with a larger group of participants, it may point to the need for hardware specifically designed to withstand the rigors of the field.

5.5.2 Validation of Experiment Design

5.5.2.1 Task Load Manipulation

In the experiment design, workload was manipulated by varying the task load (rate of incoming messages) over a block of time. The team determined whether participants subjectively experienced a significant difference in workload, as measured by their responses to the NASA TLX, by comparing the TLX scores of the high and low task load blocks in the unmitigated scenarios. Figure 53 illustrates the responses to the TLX survey after the low and high task load blocks during the unmitigated communications scenarios. During the high task load blocks, participants recorded a statistically significant increase in mental demand ($F_{1,7} = 13.4$, $p = .008$), temporal demand ($F_{1,7} = 23.5$, $p = .002$), performance ($F_{1,7} = 20.0$, $p = .003$), effort ($F_{1,7} = 25.9$, $p = .001$), and frustration ($F_{1,7} = 15.0$, $p = .006$) as compared with the low task load blocks. The only measure that did not change significantly was physical demand ($F_{1,7} = .006$, $p = .94$), which makes sense given that the scenario design (i.e., walk in a circle no matter how many messages you received) did not vary the physical demands in the two task load conditions. Thus, the designed workload manipulation was successful for the communications scenario.

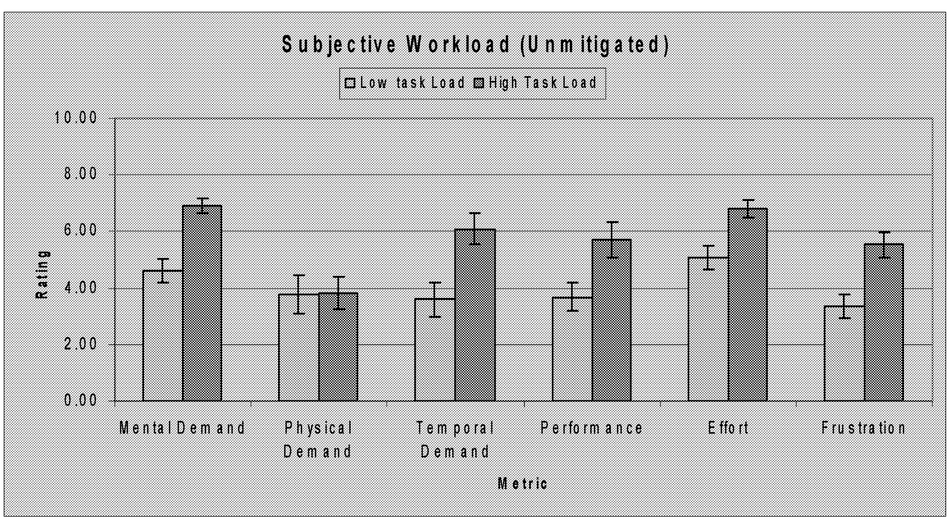


Figure 53. Subjective assessment of workload in the high and low task load blocks of the unmitigated communications scenario (bars represent standard error).

Figure 54 illustrates the responses to the TLX survey after the low and high task load blocks during the unmitigated navigation scenarios. During the high task load blocks,

participants recorded an increase (approaching significance) in most TLX measures: mental demand ($F_{1,3} = 5.12$, $p = .109$), physical demand ($F_{1,3} = 5.93$, $p = .093$), temporal demand ($F_{1,3} = 4.32$, $p = .129$), performance ($F_{1,3} = 5.56$, $p = .100$), effort ($F_{1,3} = 4.17$, $p = .134$), and frustration ($F_{1,3} = 5.84$, $p = .094$). The power of the statistical analysis was reduced since only four participants' data could be used; however, given the trends, it can be reasonably asserted that the task load manipulation was successful at placing a greater demand on the participants' cognitive resources in the high task load condition than in the low task load condition.

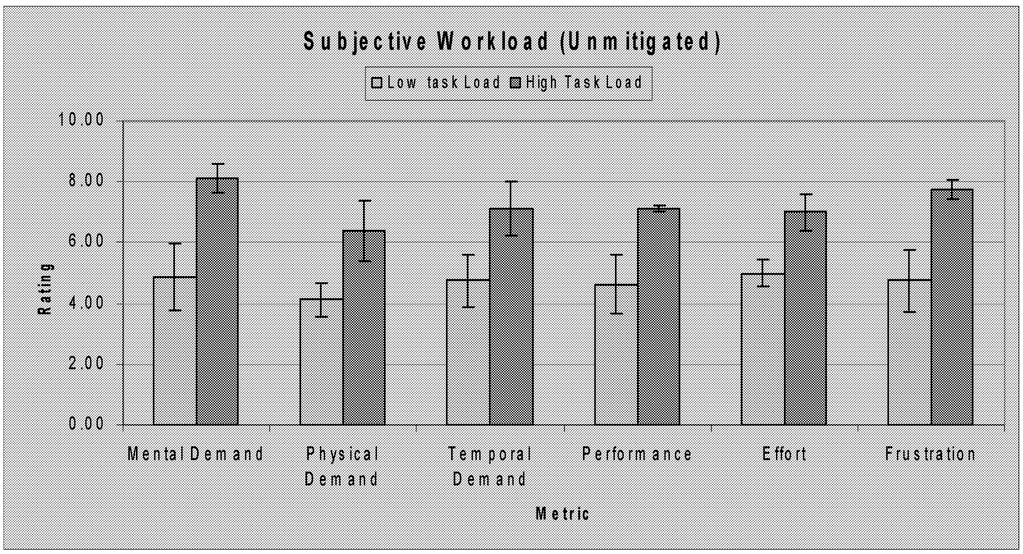


Figure 54. Subjective workload assessment during the high and low task load blocks of the unmitigated navigation scenario (bars represent standard error).

5.5.2.2 Collapsing Across Block and Scenario Version Order

The scenario was designed with several assumptions that were validated. The design had the following properties:

1. Each scenario (communications, navigation) had two isomorphic versions; thus, no systematic differences in performance should have been seen between version A and version B when all other conditions are identical. Based on this assumption, scenario version order was fixed (participants saw version A first and version B second).

2. In the communications scenario, participants experienced four task load blocks in a fixed order (high, low, high, low). Participants were trained to criteria on all tasks; thus, there should have been no "training effect" between successive high task load blocks (and similarly between low task load blocks).

Analysis of the performance data showed that these assumptions were valid since there were no statistically significant differences between the scenarios or on the order of presentation. The data were collapsed over scenario version and over task load block order.

5.5.3 Communications Scenario

This section details the data analysis results for the communication scenario.

5.5.3.1 Subjective Workload via NASA TLX

The Communications Scheduler mitigation significantly lowered the participants' subjective workload during high task load blocks of the scenario. Figure 55 illustrates the participants' subjective assessment of mental demand ($F_{1,7}$ = 28.9, p = .001), temporal demand ($F_{1,7}$ = 15.9. p = .005), performance ($F_{1,7}$ = 8.8, p = .021), effort ($F_{1,7}$ = 35.5, p<.001), and frustration ($F_{1,7}$ = 10.1, p = .016). During high task load conditions mitigation significantly improved performance. Physical demand remained unchanged ($F_{1,7}$ = 3.7, p = ..095), as was expected given the cognitive nature of the task load manipulation.

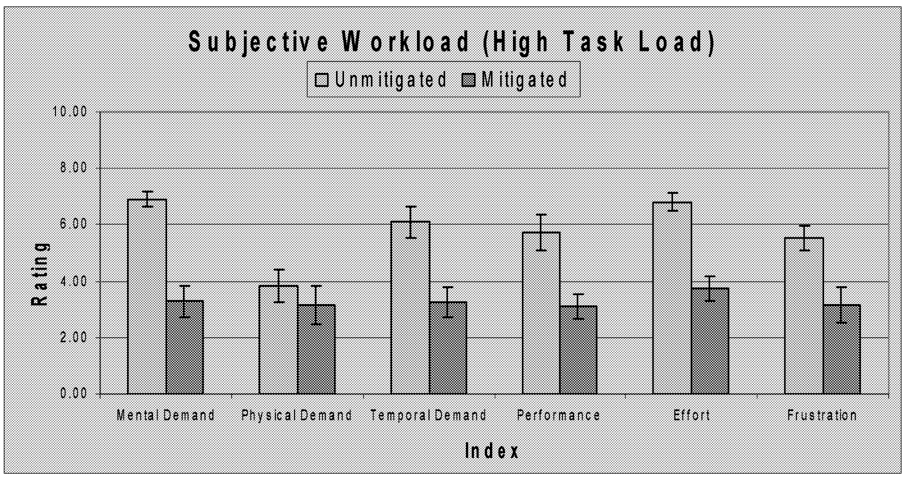


Figure 55. Subjective workload assessment during the high task load blocks of the communications scenario (bars represent standard error).

5.5.3.2 Stress

Figure 56 illustrates the subjective ratings of *performance anxiety* under the experimental conditions, where a higher rating equates to a higher stress level. Baseline stress levels averaged 2.7 for the unmitigated and 3.0 for the mitigated conditions, where the difference was not statistically significant ($F_{1,6}$ = 0.36, p = 0.569). The ARS score decreased slightly in the low task load conditions with the mitigation (2.9 unmitigated to 2.1 mitigated), but this difference was not statistically significant ($F_{1,7}$ = 2.86, p = .134). In high task load conditions, the mitigation reduced the anxiety rating from 3.6 (unmitigated) to 2.9 (mitigated), also not statistically significant ($F_{1,7}$ = 1.18, p = .314).

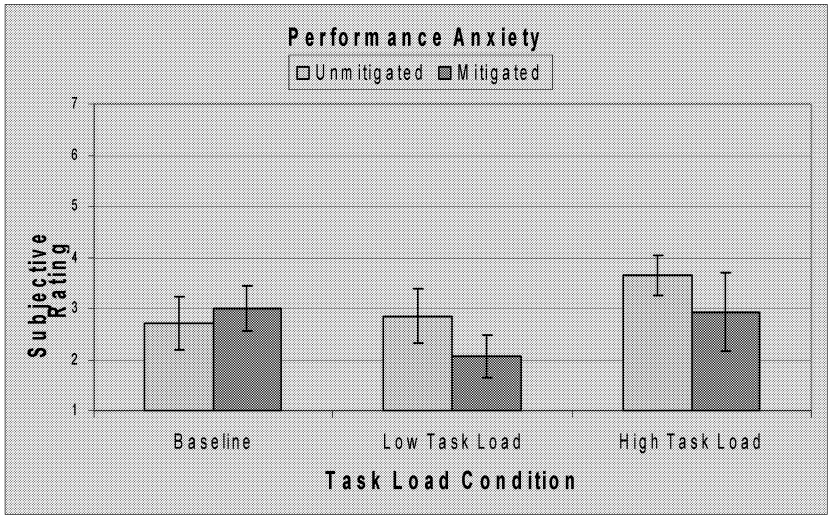


Figure 56. ARS Q1 ratings under task loads of none (baseline), low, and high for the communications scenario.

The baseline ratings for *physical anxiety* were 2.1 (unmitigated) and 2.3 (mitigated), where a higher rating equates to a higher stress level. The difference was not statistically significant ($F_{1,6} = 0.125$, $p = .736$). Mitigation lowered perceived physical stress levels in both the low and high task load conditions (see Figure 57). The stress rating in the low task load condition dropped from 2.8 (unmitigated) to 2.0 (mitigated), a statistically significant ($F_{1,7} = 11.1$, $p = .013$) difference. The ratings in the high task load condition dropped from 3.1 (unmitigated) to 2.4 (mitigated), a difference that was not statistically significant ($F_{1,7} = 3.40$, $p = .108$).

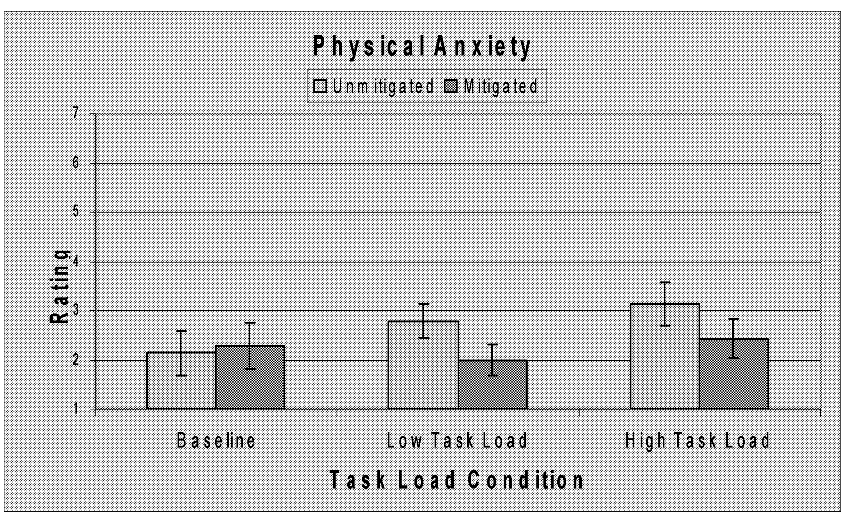


Figure 57. ARS Q2 ratings under task loads of none (baseline), low, and high for the communications scenario.

Figure 58 illustrates the subjective ratings for participants' *confidence* in their ability to perform well, where a higher rating equates to an increased confidence in performance. The baseline task load condition saw no statistically significant ($F_{1,7} = 0.77$, $p = .413$) difference in means. The participants had more confidence in their performance in both

task load conditions when the mitigation was present. Confidence increased significantly ($F_{1,7} = 7.00$, $p = .033$) from 4.3 (unmitigated) to 5.2 (mitigated) in the low task load condition and also increased significantly ($F_{1,7} = 6.45$, $p = .039$) from 3.6 (unmitigated) to 5.1 (mitigated) in the high task load condition.

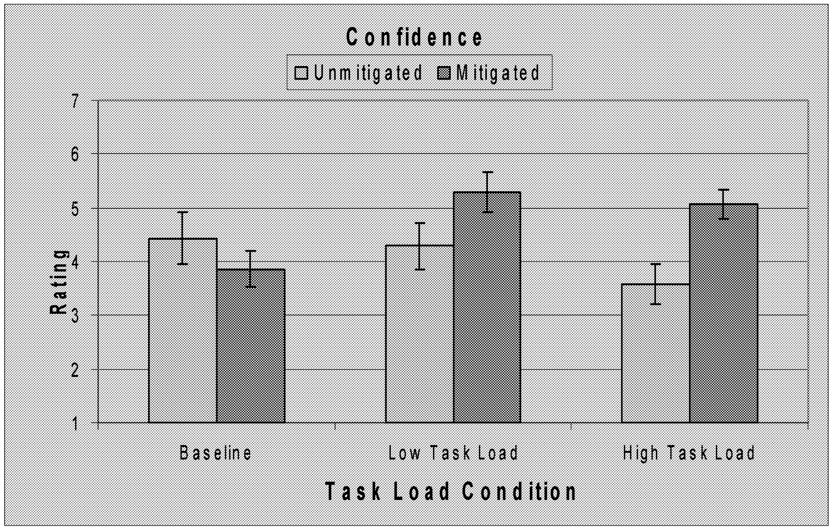


Figure 58. ARS Q3 ratings under task loads, of none (baseline), low, and high for the communications scenario.

5.5.3.3 Maintain Counts Task

Participants showed a statistically significant increase in accuracy of maintaining counts in high task load conditions when the Communications Scheduler mitigation was available (see Figure 59). Participants under high task load performed at a 67.4% accuracy when unmitigated, but their performance jumped to 95.7% accuracy when the tasks were mitigated. The effect was statistically significant ($F_{1,7} = 16.8$, $p = .005$). Under low task load, participants performed equally in both mitigation conditions ($F_{1,7} = 0.68$, $p = .440$; 83.3% (unmitigated) to 89.2% (mitigated)). This was consistent with the hypothesis that the benefits of mitigation were realized in high task load times.

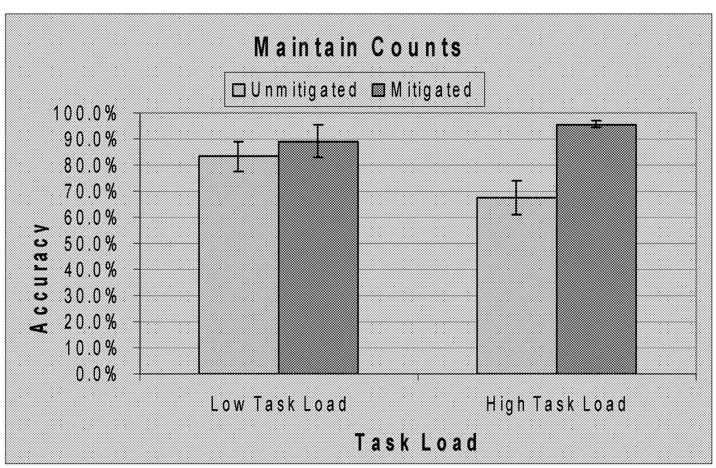


Figure 59. Accuracy of maintaining counts for the communications scenario.

5.5.3.4 *Mission Monitoring Task*

Participants showed a statistically significant increase in accuracy of mission monitoring in high task load conditions when mitigation was used (see Figure 60). Participants in high task load conditions performed at 68.2% accuracy when unmitigated, but their performance jumped to 95.8% accuracy when mitigated. The effect was statistically significant ($F_{1,7} = 18.9$, $p = .003$). In low task load conditions, participants saw a slight increase in mean performance (92.2% to 100%) with the mitigation, although this difference was not statistically significant ($F_{1,7} = 3.72$, $p = .09$). This was consistent with the hypothesis that the benefits of mitigation were realized under high task load and the resulting high-workload times.

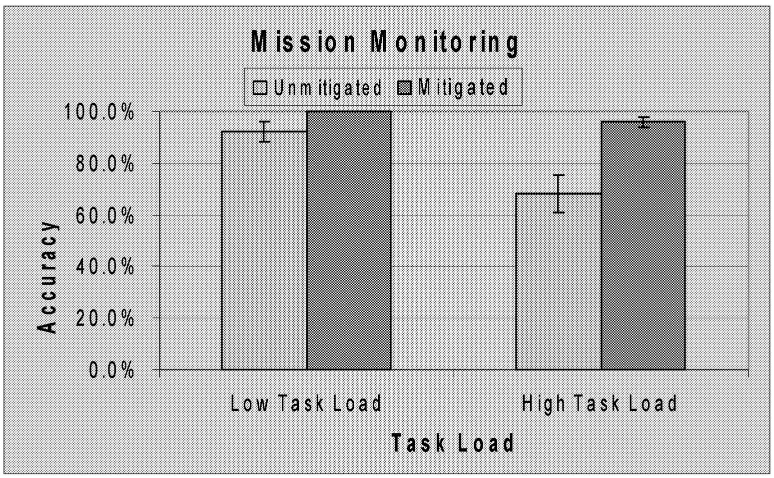


Figure 60. Accuracy of mission monitoring for the communications scenario.

5.5.3.5 *Low-Priority Message Situation Awareness*

It was hypothesized that mitigation, while providing benefit, may have had costs associated with it that made it inappropriate to leave the mitigation on all the time. To assess the possible costs of the Communications Scheduler, participants were asked SA questions at the end of each high task load block of the communications scenario. These SA questions pertained to low-priority messages that were deferred to the PDA for later review. Three questions were asked at the end of the two high task load blocks. Participants in the unmitigated scenarios scored an average of 30% (mean of 0.9 out of 3, standard error = .187) correct on the questions. The poor recall during unmitigated blocks shows that participants were fairly good at ignoring the low-priority messages. However, in the mitigated blocks, participants scored 0%, since they did not have time to review low-priority messages (see Figure 64). The effect was statistically significant ($F_{1,4} = 23.1$, $p = .009$).

Although the temporary loss of SA of low-priority messages was a potential problem, it was an acceptable cost when performance on high-priority tasks was failing due to cognitive overload. The Communications Scheduler was designed to activate when the CSA detected cognitive overload. It significantly improved performance on the high-priority tasks of maintaining counts and mission monitoring. This was an example of the benefits outweighing the costs in certain conditions. However, when participants were

able to handle all the tasks, the Communications Scheduler was not triggered, so as not to disrupt the workflow and introduce SA costs when there was no benefit to performance on high-priority tasks.

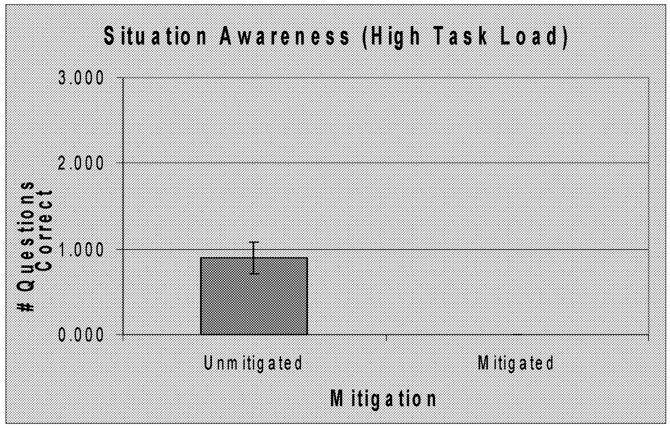


Figure 61. Situation awareness of low-priority messages in high task load blocks of the communications scenario.

5.5.3.6 Math Interruption Task

The math interruption task was used to assess attention and cognitive resources available at any given moment. Three measures are associated with the task:

- Reaction time—how quickly did the participant react to an alert? This was used to assess the attention resources available at that moment.
- Solution time—once interrupted, how much time did the participant take to correctly solve the math problem? This was used to assess the participant's focus on the math problem (i.e., how successfully was the participant interrupted?).
- Accuracy—how did the participant perform on the math problem? This was used to assess the participant's cognitive ability or spare cognitive resources to perform the task.

Due to data logging issues, only four of eight participants' data were recorded for the low task load condition of the math task. Seven of eight participants' data were used in the analysis of the high task load condition for the math interruption task.

Participants responded to the interruption alert much more quickly in the low task load condition than the high task load condition, as expected. In the low task load condition, mitigation slightly decreased reaction time from 4.9 seconds (unmitigated) to 3.7 seconds (mitigated), although not significantly ($F_{1,3} = 1.85$, $p = .267$). In the high task load condition, where benefits of mitigation were expected, reaction time was faster under mitigation by almost 5 seconds going from 8.6 seconds (unmitigated) to 3.8 seconds (mitigated), as illustrated in Figure 62. The difference approached significant ($F_{1,6} = 4.8$, $p = .070$).

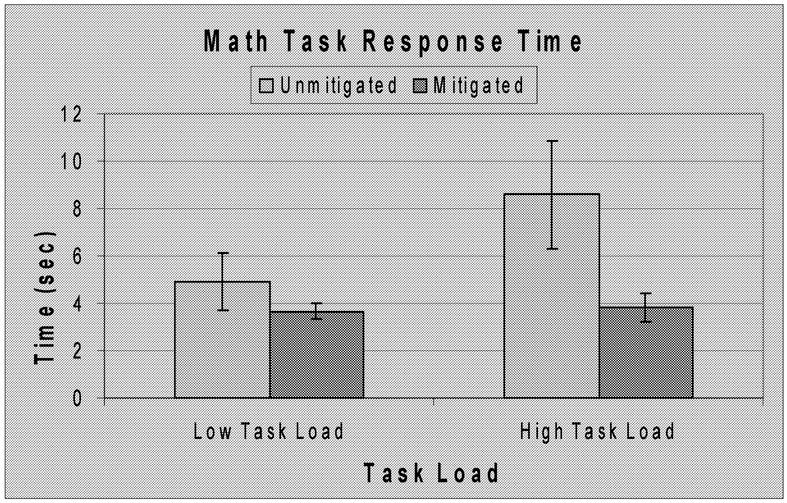


Figure 62. Reaction time for the math interruption task in the communications scenario.

Once interrupted, the participant's time to actually solve the math problem did not vary significantly across any of the experimental conditions (see Figure 63). In low task load ($F_{1,3} = 0.48$, $p = .536$) conditions, participants' time to solve the math problem was 6.9 seconds (unmitigated) and 6.3 seconds (mitigated). In high task load ($F_{1,6} = 0.898$, $p = .380$) conditions, participants solved the math problem in the same amount of time: 6.5 seconds (unmitigated) to 6.9 seconds (mitigated). The data suggest that once the participants were interrupted, their entire attention was focused on solving the math problem, with little variation between experimental conditions.

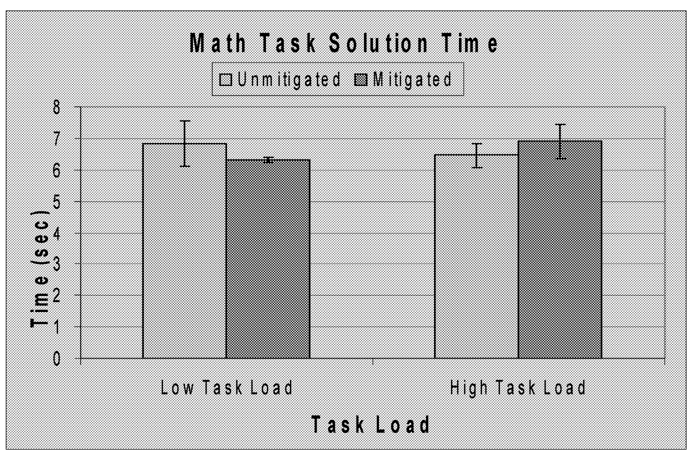


Figure 63. Solution time for the math interruption task in the communications scenario.

Participants' accuracy in solving the math problems did not vary significantly between the task load conditions or between the mitigation conditions (see Figure 64). In low task load conditions, the accuracy was the same in both conditions: 76.0% (unmitigated) and 75.0% (mitigated), ($F_{1,3} = .005$, $p = .949$). In high task load conditions ($F_{1,6} = 0.013$, $p = .914$), the accuracy was again the same in both conditions: 78.6% (unmitigated) and 77.6% (mitigated). The data suggested that once participants were interrupted and solving the task, they dedicated the necessary amount of resources to solving the math problems accurately.

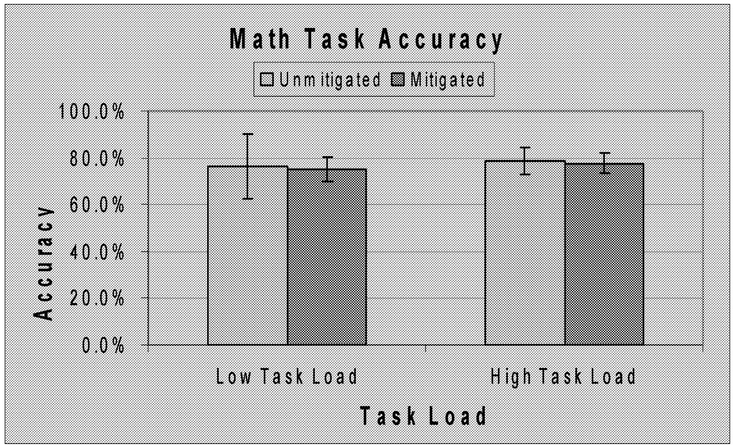


Figure 64. Accuracy for the math interruption task in the communications scenario.

5.5.4 Navigation Scenario

This section details the data analysis results for the navigation scenario.

5.5.4.1 Subjective Workload Assessment

The Tactile Navigation Cueing System lowered the subjective workload of the participant during high task load blocks of the scenario. Participants' subjective assessment of mental demand ($F_{1,3} = 3.39$, $p = .163$), physical demand ($F_{1,3} = 1.57$, $p = .299$), temporal demand ($F_{1,3} = 6.04$. $p = .091$), performance ($F_{1,3} = 3.39$, $p = .163$), effort ($F_{1,3} = 1.5$, $p = .308$), and frustration ($F_{1,3} = 3.35$, $p = .165$) was improved by mitigation over the unmitigated trials during high task load (see Figure 65). Although none of the measurable differences reached the significance threshold of $p<.05$, the data indicated that the work-load means were consistently reduced across each category in the mitigated condition. Note that all the scores were high, indicating that this scenario was very taxing overall.

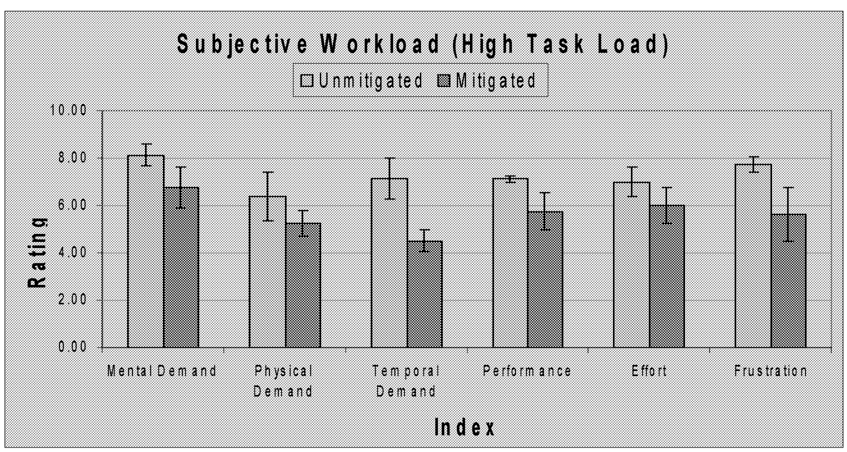


Figure 65. Subjective workload assessment in high task load conditions for the navigation scenario.

5.5.4.2 Stress Assessment

Figure 66 illustrates the subjective ratings of *performance anxiety* under the experimental conditions, where a higher rating equates to a higher anxiety level. Baseline stress levels

averaged 3.0 for unmitigated and 2.0 for mitigated, where the difference was not statistically significant ($F_{1,1} = 49$, $p = .090$). Ratings in the low task load conditions were similar in both mitigation conditions (3.5 unmitigated, 3.0 mitigated, $F_{1,3} = 3.0$, $p = .182$). Performance anxiety ratings in the high task load conditions, however, were reduced slightly in the mitigated conditions, producing a result approaching significance (unmitigated 5.25, mitigated 3.5; $F_{1,3} = 7.74$, $p = .069$). In other words, the mitigation was only effective at reducing performance-anxiety-related stress in the high task load mitigated conditions as compared to low task load conditions.

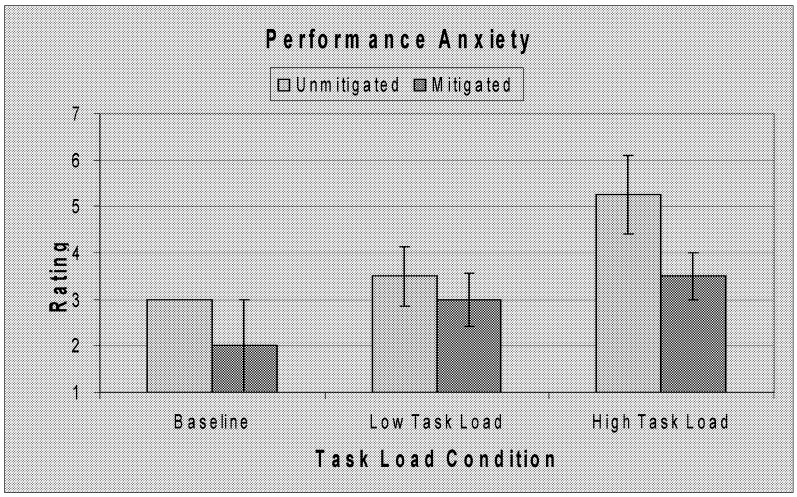


Figure 66. Nav. scenario ARS Q1 ratings for task loads of none (baseline), low, and high.

The baseline ratings for *physical anxiety* were 2.0 (unmitigated) and 2.5 (mitigated), where a higher rating equates to a higher stress level ($F_{1,1} = 4.0$, $p = .295$). Mitigation lowered stress levels in both the low and high task load conditions, although none of the differences were statistically significant (see Figure 67). Ratings in the low task load conditions ($F_{1,3} = 6.00$, $p = .092$) were similar with ratings of 3.75 (unmitigated) and 2.75 (mitigated). High task load ratings were also equal ($F_{1,3} = 3.00$, $p = .182$) in both conditions—4.75 (unmitigated) and 4.25 (mitigated).

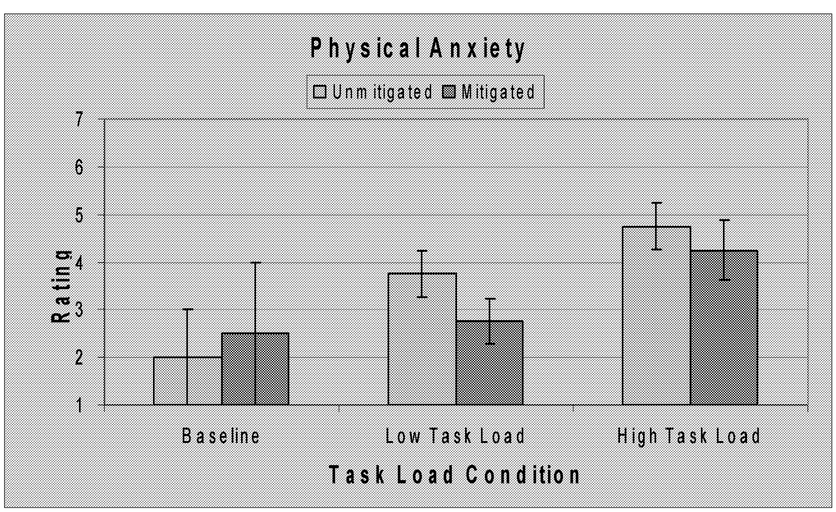


Figure 67. Nav. scenario ARS Q2 ratings for task loads of none (baseline), low, and high.

The baseline ratings for *confidence* were 4.0 (unmitigated) and 4.5 (mitigated), where a higher rating equates to a higher confidence level, and the difference was not statistically significant ($F_{1,1} = 25.0$, $p = .126$) (see Figure 68). Confidence in performance fell in each task load condition as compared with the initial baseline condition. The availability of the mitigation increased confidence over the unmitigated case in each task load condition, although none of the differences were significant. Low task load confidence was generally equivalent in both mitigation conditions (4.75 unmitigated, 5.0 mitigated, $F_{1,3} = 0.273$, $p = .638$). High task load confidence increased from 2.5 (unmitigated) to 4.5 (mitigated) ($F_{1,3} = 6.00$, $p = .092$).

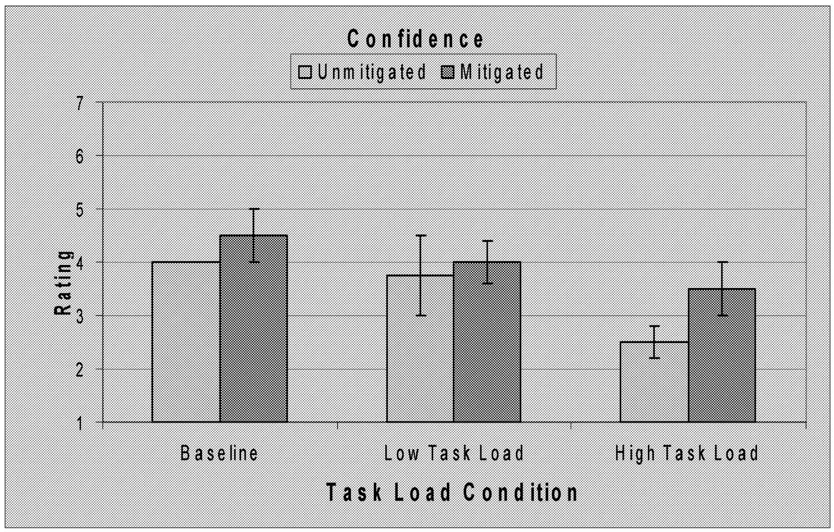


Figure 68. Nav. scenario ARS Q3 ratings for task loads of none (baseline), low, and high.

Overall, mitigation improved participants' stress levels, most notably under the high task load conditions.

5.5.4.3 Maintain Counts Task

Although the navigation scenario shared some tasks with the communications scenario, scores were uniformly lower in the navigation scenario due to the nature of the additional visual and navigation tasks imposed. Participants showed an increase in accuracy of maintaining counts in high task load blocks when tactile cueing mitigation was available (see Figure 69). Note that the mitigation did not directly influence the task in this case, but its presence freed up cognitive resources that were devoted to the navigation task. However, in low task load blocks, mitigation actually reduced performance. Participants in high task load conditions performed at equivalent levels, having 29.9% accuracy when unmitigated and 35.1% accuracy when mitigated. There was no difference in these accuracies ($F_{1,5} = 0.42$, $p = .547$). In the low task load blocks, participants saw a statistically significant ($F_{1,5} = 7.71$, $p = .039$) decrement in performance in the mitigated condition, 26.2% accuracy, while unmitigated accuracy was 43.3%. This was consistent with the hypothesis that the application of mitigation resulted in a cost to performance if not appropriately applied to the situation. For example, it was possible that the mitigation (tactile buzzing) proved to be a distraction to competing tasks when walking in a straight line in the low task load block.

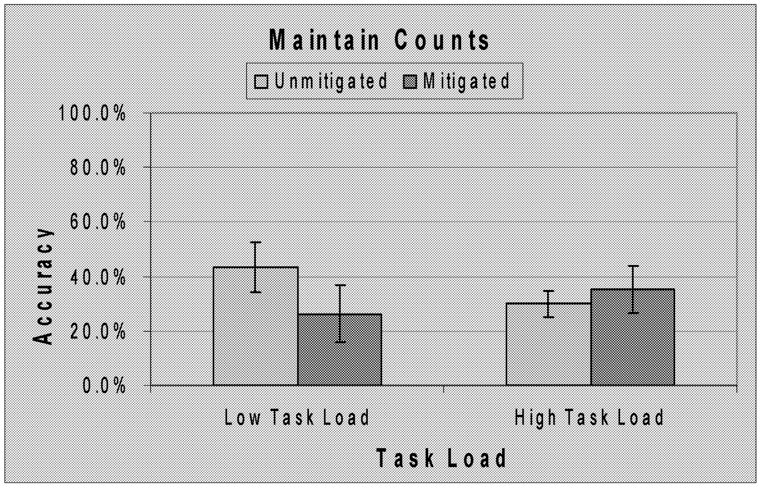


Figure 69. Maintain counts accuracy for the navigation scenario.

5.5.4.4 Mission Monitoring Task

Participants performed equivalently on the mission monitoring task in high task load conditions when mitigation was available (see Figure 70). Performance in high task load blocks was similar in the unmitigated (35.2%) and mitigated (49.3%) conditions. Likewise, in low task load blocks, the accuracy was similar, with 83.3% in the unmitigated condition and 100% in mitigated. While neither the high task load ($F_{1,5}$ = .917, p = .382) nor the low task load ($F_{1,5}$ = 1.00, p = .363) differences were significant, the results trended in the direction of the mitigation, positively influencing performance.

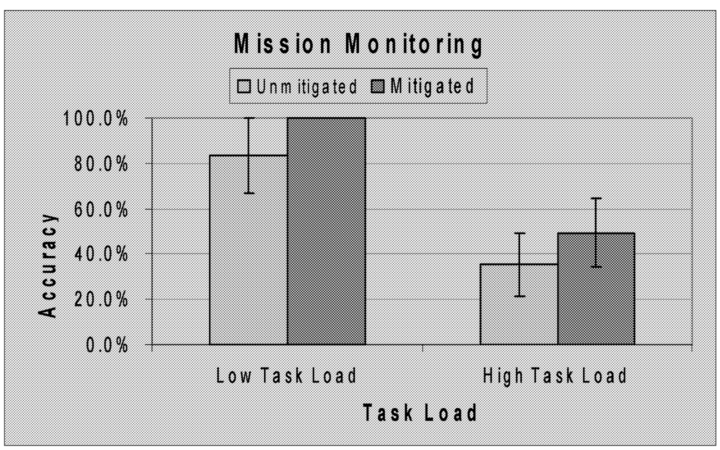


Figure 70. Mission monitoring accuracy for the navigation scenario.

5.5.4.5 Math Interruption Task

The math interruption task was used to assess attention and cognitive resources available at any given moment. Due to data logging issues, only two of six participants' data were recorded for the low task load condition of the math task. Data for all six participants were used in the analysis of the high task load condition for the math task.

Participants responded to the interruption alert much more quickly in the low task load conditions, as expected. In the low task load conditions, the mitigation actually slightly

increased reaction time from 3.3 seconds (unmitigated) to 5.2 seconds (mitigated), although not significantly ($F_{1,1} = 1.00$, $p = .363$). In the high task load conditions, where benefits of mitigation were expected, reaction time was reduced under mitigation, from 22.0 seconds (unmitigated) to 9.1 seconds (mitigated), as illustrated in Figure 71. Although the difference was not statistically significant ($F_{1,5} = 1.20$, $p = .324$), the reduction was in the positive direction.

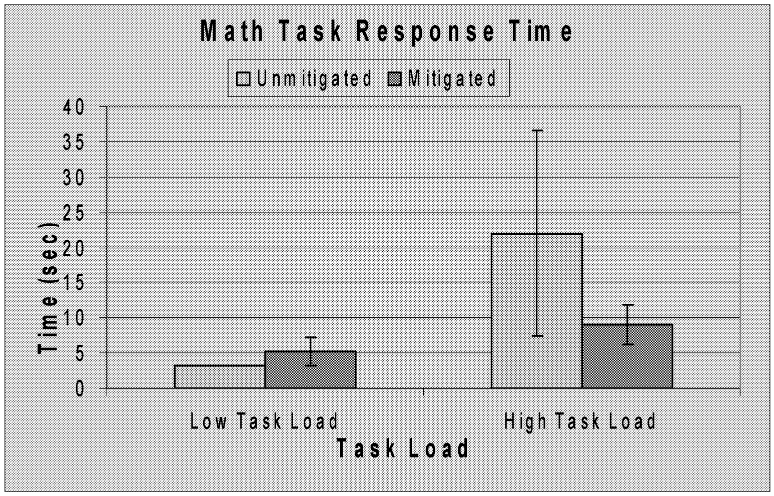


Figure 71. Reaction time for the math interruption task in the navigation scenario.

Once the participants were interrupted, their time to actually solve the math problem presented to them did not vary significantly across any of the experimental conditions (see Figure 72). In low task load conditions, participants solved the math problem faster in the mitigated case (4.7 seconds) as compared with the unmitigated (6.4 seconds), although this was not statistically significant ($F_{1,1} = 0.963$, $p = .506$). In high task load conditions, participants' performance was unaffected by the mitigation ($F_{1,5} = 0.184$, $p = .686$): 6.9 seconds (unmitigated) to 6.6 seconds (mitigated). The data suggested that once the participants were interrupted, their entire attention was focused on solving the math problem.

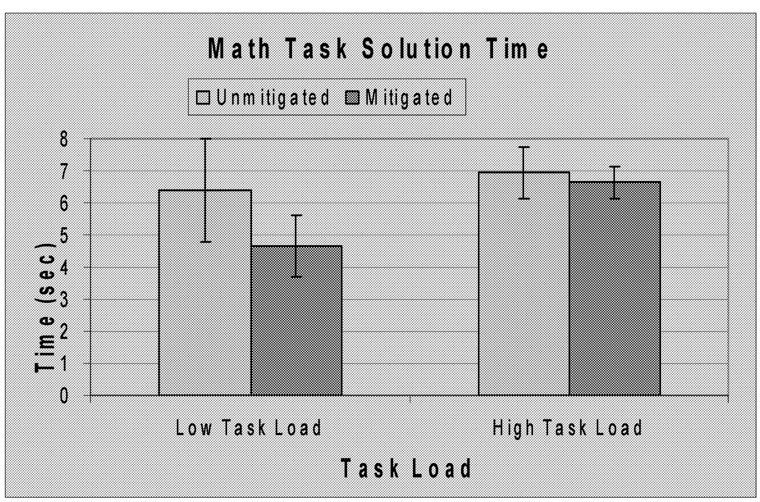


Figure 72. Solution time for the math interruption task in the navigation scenario.

Participants' accuracy in solving the math problems was considerably reduced in the high task load as compared with the low task load conditions (see Figure 73). In addition, in each task load condition, mitigation increased participants' accuracy. In low task load, accuracy was similar, with 83.3% accuracy in the unmitigated condition and 100% in the mitigated case ($F_{1,1} = 6.84$, $p = .232$). In high task load, accuracy was significantly ($F_{1,5} = 7.26$, $p = .043$) increased from 47.0% (unmitigated) to 68.0% (mitigated). Again, the principal benefits of the mitigation were seen in the high task load conditions.

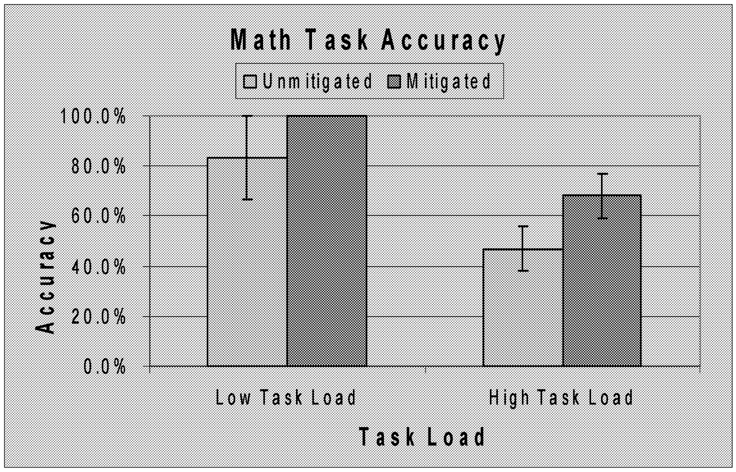


Figure 73. Accuracy for the math interruption task in the navigation scenario.

5.5.4.6 Navigation Task

The mitigation directly targeted performance on the navigation task. Figure 74 illustrates the composite runtime for all experimental conditions. The differences between composite runtime in low task load vs. high task load were due solely to the fact that the path in the high task load block was considerably longer than the path in the low task load block.

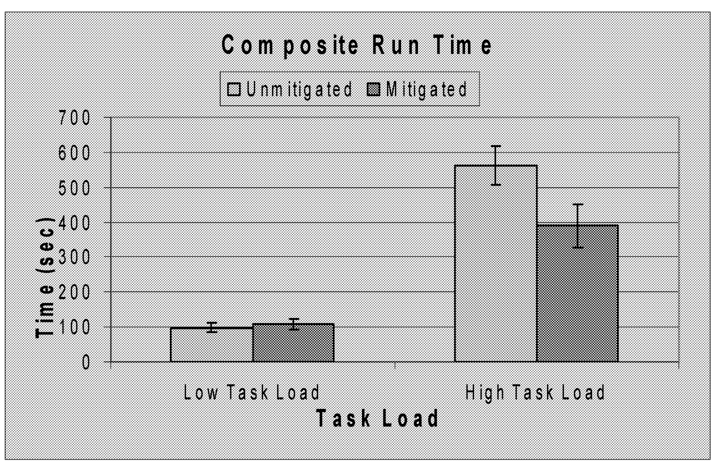


Figure 74. Composite runtime for the navigation scenario.

Mitigation showed no statistically significant ($F_{1,5} = .265$, $p = .629$) effect on composite runtime in the low task load conditions: 98.3 seconds (unmitigated) vs. 107.7 seconds (mitigated). In high task load, however, the Tactile Navigation Cueing System enabled a

statistically significant ($F_{1,5}$ = 8.69, p = .032) reduction in composite runtime, from 562.1 seconds (unmitigated) to 389.2 seconds (mitigated). Thus, in high task load conditions, participants were able to navigate to the objective more quickly and more safely.

5.5.4.7 Visual Search for IEDs

The results of participants' search for IEDs are illustrated in Figure 75. Participant performance did not vary significantly either under task load or mitigation conditions. In low task load, participants' search accuracy was not significantly ($F_{1,5}$ = 0.29, p = .611) different: 41.7% (unmitigated) and 33.3% (mitigated). In high task load, participants' search accuracy was unchanged from 50.0% in both the unmitigated and mitigated cases ($F_{1,5}$ = 0.00, p = 1.0). One might reasonably expect that the mitigation would free up resources to scan the environment. With no mitigation, the participants were forced to scan the environment looking for visual cues, where they would also detect IEDs. Unfortunately, the data are inconclusive.

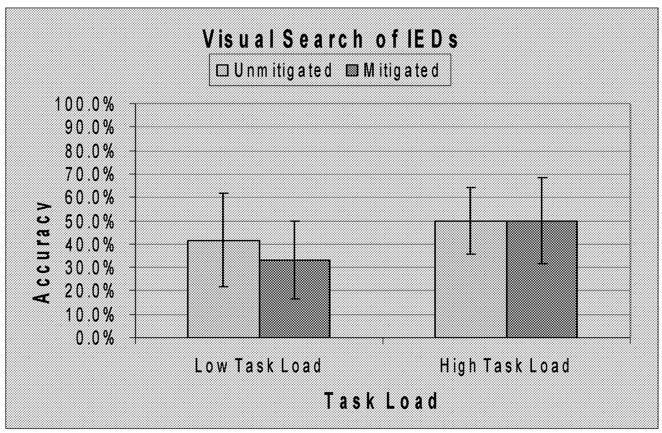


Figure 75. Visual search for IEDs in the navigation scenario.

5.5.4.8 Path Situation Awareness Assessment

Participants were asked to draw, on a blank map with only landmarks, the path they had just traversed. It was hypothesized that participants in the mitigated case, where they were being guided through the minefield by the Tactile Navigation Cueing System, might have suffered some loss of SA of their surroundings. However, this hypothesis was not supported. As shown in Figure 76 the mean difference in the participants' drawn paths from their true path was similar in the mitigated and unmitigated conditions. For high task load, path difference in the unmitigated condition was 2212 m^2, whereas path difference in the mitigated condition was 2342 m^2 which was not a statistically significant difference ($F_{1,4}$ = .095, p = .773). Likewise, the low task load condition showed no statistically significant difference ($F_{1,4}$ = 5.09, p = .087) in the means between the unmitigated (342 m^2) and the mitigated (255 m^2) conditions. Note that the large difference in path deviation between high and low task loads was due to the differing length of the path in those blocks.

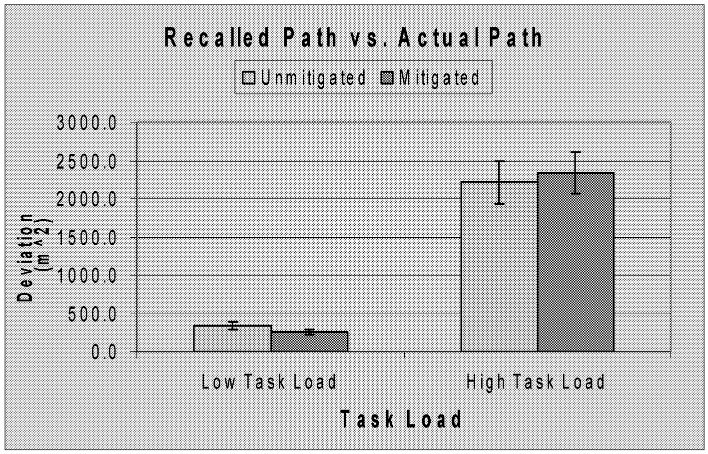


Figure 76. Path situation awareness for the navigation scenario.

5.5.5 Cost/Benefit Analysis

The Spring CVE contained two scenarios, each with a host of metrics under two experimental conditions: task load and mitigation. The previous sections discussed each metric in detail and context, discussing the benefits and costs of each in relation to the mitigations. For an overview of where mitigations produce benefits or induce costs, see Table 35.

Table 35. Summary of the benefits/costs of mitigation.

Measure/Task	Communications Scenario		Navigation Scenario	
	Low Task Load	High Task Load	Low Task Load	High Task Load
Subjective Workload	+	+		+
Performance Anxiety				+
Physical Anxiety	+			
Confidence	+	+		+
Maintain Counts		+	-	
Mission Monitoring		+		
Low-Priority Message SA		-	n/a	n/a
Interruption Task Reaction Time		+		
Interruption Task Solve Time				
Interruption Task Accuracy				+
Visual Scan for IEDs	n/a	n/a		
Composite Runtime	n/a	n/a		+
Paths Situation Awareness	n/a	n/a		

Key:	Performance Improvement
+	Statistically significant improvement
–	Statistically significant decrement

In the communications scenario, the Communications Scheduler successfully improved performance of the target tasks of maintaining counts and mission monitoring during high-workload periods. In addition, participants had more attentional resources to react to

interruptions. However, SA of deferred low-priority messages suffered. Specifically, the communications scenario resulted in the following conclusions:

Task manipulation successful:

- Subjective workload assessment agreed with task load manipulation. That is, participants reported higher subjective workload in the high task load condition than in the low task load condition.
- Response time in the math task was faster for low task load conditions than for high task load conditions. That is, participants had more cognitive resources available in low task load than in high task load conditions.
- Conclusion: the CVE was able to create periods of high workload task load in the participant via the task load manipulation.

Mitigation produced:

- Lower reported workload in both high and low task load conditions (benefit)
- Higher confidence in participants during both high and low task load conditions (benefit)
- Performance increased in the mitigated task of maintaining counts in high task load condition (benefit)
- Performance increased in the competing task of mission monitoring in high task load condition (benefit)
- Decreased SA for low-priority messages in high task load condition (cost)
- Decreased response time in the math task in high task load conditions (benefit)

In the navigation scenario, the cost/benefit tradeoffs of mitigation were even more pronounced. The Tactile Navigation Cueing System, by relieving participants of the cognitively challenging task of navigating through an unfamiliar area, resulted in the improvement of almost all tasks in the high task load condition. However, when the Tactile Navigation Cueing System was invoked during low task load periods, it was so distracting that almost all tasks suffered as a result. Specifically, the following conclusions can be drawn from the navigation scenario:

Task manipulation successful:

- Subjective workload assessment agreed with task load manipulation, i.e., participants reported higher subjective workload in the high task load condition than in the low task load condition.

Mitigation produced:

- Marginal reduction in subjective workload in the high task load condition (benefit)
- Higher confidence in participants during the high task load condition (benefit)
- Marginal reduction in performance anxiety during the high task load condition (benefit)

- Decreased performance in the competing task of maintaining counts during the low task load period (cost)
- Higher accuracy in the competing math task during the high task load periods (benefit)
- Decreased runtime for the mitigated task of navigate to objective during the high task load period (benefit)

Thus, it can be concluded that both mitigations were most effective when used during high task load periods. Costs involved when mitigations were inappropriately invoked during low task load periods resulted in significant performance degradation. Closing the loop with an accurate assessment of cognitive state, in order to appropriately trigger mitigation, was vital for the mitigations to prove effective in real operational settings.

5.6 Phase 3 Joint Distributed Freeplay Event

5.6.1 Overview

The Honeywell AugCog team participated with the Aberdeen Test Center (ATC) in a Joint Distributed Freeplay Event (JDFE) at Mulberry Point at Aberdeen Proving Ground, Maryland. Over a two-week period from August 23 through September 1, 2005, eight separate scenarios were run involving live, virtual, and distributed assets. The premise of the scenario was a joint personnel recovery mission in which a downed pilot was captured by enemy insurgents and a rescue mission was planned and executed. The AugCog team outfitted the Joint Task Force (JTF) Commander with a six-channel wireless EEG cap manufactured by ABM that was integrated into Honeywell's information architecture. Additional data collected by ATC on the commander included an ECG belt for heart rate (from Quasar) and core body temperature. A full set of data was collected on three separate days on three different Army personnel playing the role of the commander. The role of the JTF Commander primarily involved direct communications with the JTF staff to gather intelligence regarding movements of opposing force (OPFOR) and communications with the friendly (blue) force (BLUFOR) squad leader leading the recovery mission in the field. The commander had direct access to a video feed from an unmanned air vehicle (UAV) flying over the compound, a Commander's Digital Assistant (CDA), and a map of the village and compound area. Based on the incoming information, the commander directed the BLUFOR operations to recover the pilot, call in a medevac, if necessary, and extract the Soldiers from the area.

Participation in this exercise addressed the following three purposes:

- Sensor and computational system, deployment into Army-relevant environment
- Post hoc classification of cognitive states induced by Army exercise
- Preparation for AugCog Phase IV field exercises.

5.6.2 Operational Scenario

The 2005 JDFE simulated squad operations in hostile territory. BLUFOR was tasked with the return of a downed pilot who was being held hostage by an insurgent group, referred to as OPFOR, in an urban environment. Depending on that day's scenario,

OPFOR assumed one of the following fighting dispositions: withdraw, engage in a short fight and then surrender, or fight to the death. In addition to organic assets, such as unmanned ground and air vehicles, the BLUFOR commander, code-named Blue-6, could task other theater assets such a simulated AC 130 Gunship.

BLUFOR squad members in the field were outfitted with LaserTag with Simunitions (soap bullets that hurt but do not injure), GPS-based wireless BLUFOR tracking, and electrocardiogram (ECG) sensor and data collection unit. The BLUFOR commander was stationed in a simulated mobile command center and was outfitted with the AugCog ABM EEG sensor headset, an ECG sensor and data collection unit, a CDA, UAV video-feed monitors, and a squad radio.

The exercise was conducted at Mulberry Point within the Aberdeen Proving Ground. This is a configurable military operations in urban terrain (MOUT) site that had extensive video sensing capability for monitoring and after-action reviews. ATC personnel simulated all smoke and explosive munitions missions in a controlled and safe manner.

5.6.3 Operational Tasks

The BLUFOR commander (code-named Blue-6) was stationed in a simulated mobile command center where he was seated in front of UAV with video-feed monitors. Once in position, Blue-6 radioed the squad leader in the field, code-named Blue-whiskey, to provide the mission brief. During this time, the squad took a simulated helicopter flight to the theater and Blue-6 coordinated pre-mission activities with the UAV operators, code-named Eagle, and simulated theatre assets, code-named Patriot. Blue-6 also coordinated with the exercise operations controllers to keep apprised of simulated and organic asset status, OPFOR readiness, and other exercise conditions. This pre-mission typically lasted 15-25 minutes.

Once Blue-6 received the go-ahead from operations control, he radioed Blue-whiskey to begin execution of the search and rescue mission. Simultaneously, the observer/controller (O/C) informed Blue-6 that the UAV was ready for deployment, and Blue-6 determined where to focus Eagle's reconnaissance video cameras in order to locate insurgents and the downed pilot. After carefully reviewing the video feeds on his monitors, Blue-6 either located insurgents and/or the pilot or redirected the UAV to monitor another location. Upon identifying the location of insurgents or the pilot, Blue-6 would communicate this information to Blue-whiskey over the squad radio. This pre-assault phase typically lasted 20-40 minutes.

Depending on the OPFOR disposition, Blue-6 might call for an AC 130 strike or work with Blue-whiskey to formulate an assault plan on the enemy location. If the assault required the squad to move through an open area, Blue-6 would order a smoke mission to obscure the movements of the squad. Blue-6 communicated the smoke mission to the O/C, who then radioed pyrotechnicians in the field who would ignite the smoke grenades.

Once Blue-whiskey initiated the assault phase, Blue-6 would be unable to raise his squad on the radio. Typically, Blue-6 would sit back and wait to receive an update from Blue-whiskey. During the assault, bursts of gunfire could be heard from the MOUT site located approximately 200 meters away. This assault phase typically lasted 10-20 minutes.

5.6.4 Participants

Three U.S. Army sergeants were assigned the role of BLUFOR commander during this exercise.

5.6.5 Sensor System

During this exercise, Honeywell outfitted the BLUFOR commander with ABM's EEG sensor headset. The sensor headset acquired six channels of EEG using a bipolar montage. Differential EEG are sampled from bipolar channels CzPOz, FzPOz, F3Cz, F3F4, FzC3, C3C4 at 256 samples per second with a bandpass from 0.5 and 65 Hz (at 3-dB attenuation) obtained digitally with Sigma-Delta A/D converters. Data were transmitted across a Bluetooth radio frequency (RF) link to the collection laptop via an RS-232 interface. Quantification of the EEG in real time was achieved using signal analysis techniques to identify and decontaminate eye blinks and to identify and reject data points contaminated with electromyography (EMG), amplifier saturation, and/or excursions due to movement artifacts (see Berka, Levendowski, Cvetinovic, Petrovic, et al., 2004, for a detailed description of the artifact decontamination procedures). Decontaminated EEG was then segmented into overlapping 256-data-point windows called overlays. An epoch consisted of three consecutive overlays. Fast-Fourier transform was applied to each overlay of the decontaminated EEG signal multiplied by the Kaiser window ($\alpha = 6.0$) to compute the power spectral densities (PSDs). The PSD values were adjusted to take into account zero values inserted for artifact-contaminated data points. The PSD between 70 and 128 Hz was used to detect EMG artifacts. Overlays with excessive EMG artifacts ("EMG") or with fewer than 128 data points ("missing data") were rejected. The remaining overlays were averaged to derive PSDs for each epoch with a 50% overlapping window. Epochs with two or more overlays with EMG or missing data were classified as invalid. For each channel, PSD values were derived for each 1-Hz bin ("bin") from 3 to 40 Hz and the total PSD from 3 to 40 Hz ("band"). "Relative power" variables were also computed for each channel and bin using the formula ("total band power/total bin power").

During collection, the information architecture was summed across 1-Hz bins to calculate and log relative power in the following EEG bands: theta, alpha, beta, high beta, and gamma. Relative power for each of the five bands for the six differential channels (CzPOz, FzPOz, F3Cz, F3F4, FzC3, C3C4) yielded 30 features to be investigated in post hoc analyses.

5.6.6 JDFE Analysis

Using the variations in the cognitive workload required of the scenario, the Honeywell AugCog team evaluated the classification techniques previously used in the laboratory and field tests to classify performance.

5.6.6.1 Task Characterization

After observing the BLUFOR commander for two days, Honeywell personnel identified the following salient tasks of their mission:

- Communicating: The BLUFOR commander maintained awareness and initiated mission actions via communicating over his squad radio. He communicated with his squad leader in the field, joint fire assets, and UAV operators.

- UAV monitoring: During this exercise, the commander tasked a prototype UAV with video surveillance to survey the mission area. He closely monitored the video feed on a small monitor in order to identify enemy locations and to locate the downed pilot.

- Interaction with CDA: The commander interacted with his CDA, which is a ruggedized PDA, to send mission directives and text messages to the squad in the field.

- Working with paper map: The commander used a paper map to update and maintain his situation awareness of the evolving mission.

- Interaction with operation control: During this complicated exercise, observer/controllers (O/Cs) worked to ensure that all elements were coordinated to maintain operational realism and ensure safety of the participants. For example, O/C monitored the status of the UAV and updated the commanders throughout the exercise; furthermore, O/C ordered smoke and munitions missions for the pyrotechnicians to execute. The O/C kept the commander up-to-date regarding the overall exercise timing, and this required frequent interaction.

- Waiting: At different points in the mission, the commander tasked his squad in the field to execute some task and then would typically wait to receive feedback from the squad leader regarding its execution. During this waiting period, the squad leader did not communicate, so radio communications decreased dramatically.

During the subsequent three collection days, an observer monitored the exercise and recorded which tasks occurred for each 15-second time block. For example, if the commander was listening on the radio while monitoring the UAV video feed, the observer recorded that those two tasks occurred during the time block in question. Subsequently, another observer reviewed the video logs of the exercise and recorded the task profiles. Inter-rater reliability analyses were conducted to ensure that there was substantial agreement between the two observers.

For the purpose of post-hoc cognitive state classification, low- and high-workload periods were operationalized as follows:

- Low workload: a 15-second time period during which the commander was not doing any of the identified tasks except for waiting.

- High workload: a 15-second time period during which the commander executed at least two of the identified tasks, not including waiting. The premise is that these periods required either multitasking or task-switching behaviors.

5.6.6.2 Cognitive State Classification

The classification approach was evaluated using a leave-one-out training and testing procedure. Given *n* data samples, the data is split into two parts: a training set consisting of *n-1* samples and a testing sample. This procedure is repeated *n* times with a different sample being chosen as the test sample each time. The average classification error over *n* trials provides an estimate of a classifier's error rate. Leave-one-out validation has been shown to be an approximately unbiased estimate of a classifier's generalization error (Efron, 1983). Such an approach does not systematically over- or underestimate the quantity being estimated. Leave-one-out testing is a computationally expensive procedure and is only practical with relatively small data sets. The classification results for each participant are shown in Table 36. Accuracy for each participant is calculated by averaging the diagonals in the confusion matrix. Average classification accuracy across the three participants was 73.6% with a range of 65.9% to 78.2%. These results clearly demonstrated that the classification approach developed as part of Honeywell's AugCog effort provided the basis for robust classification in operationally relevant task environments.

Table 36. Classification results from three participants in the JDFE.

Participant 7777	Actual Low	Actual High
Classification Low	73.53 %	26.47 %
Classification High	20.25 %	79.75 %
Accuracy	**76.64%**	
Participant 8888	**Actual Low**	**Actual High**
Classification Low	81.94 %	18.06 %
Classification High	25.45 %	74.55 %
Accuracy	**78.24%**	
Participant 9999	**Actual Low**	**Actual High**
Classification Low	64.11 %	35.89 %
Classification High	32.22 %	67.78 %
Accuracy	**65.94 %**	

Although these results were promising, the lack of experiment control introduced several caveats that will have to be addressed in future work. First, the task load labels were subjectively assigned by independent raters on the basis of behavioral observations. As such, these ratings only provide an indirect estimate of a participant's workload. Second, workload ratings assigned by raters were likely to be influenced by the verbal and behavioral expressiveness of a participant. Third, the classification evaluation was limited to a single session. The ability of the classifier to generalize broadly over large temporal windows in operationally relevant contexts remains to be established.

5.7 Phase 3 Discussion

Phase 3 culminated with the demonstration of a mobile AugCog system in an operational context. The Phase 3 Spring CVE evaluated a fully mobile CLIP and demonstrated a measurable improvement in workload, performance, and confidence when mitigations, triggered by a real-time assessment of cognitive state, assisted participants in managing task load. Subsequently, through the JDFE experience, Honeywell was able to

demonstrate cognitive state classification in an operational domain, with real Soldiers as participants, thus taking the first step toward a full evaluation in a realistic Army operational setting.

6 Augmented Cognition Program Phase 4

6.1 *Phase 4 Introduction*

6.1.1 Phase 4 Research Team

The Honeywell Augmented Cognition (AugCog) team in Phase 4 consisted of the collaborative efforts of Honeywell Laboratories, Advanced Brain Monitoring, Inc. (ABM), and Oregon Health and Sciences University. AugCog Test Event (ACTE) was the collaborative effort of the Honeywell team, the Battle Lab Integration Team (BLIT), the Aberdeen Test Center (ATC), USARIEM, Hidalgo Inc., the Army Research Lab (ARL) Human Research and Engineering Directorate (HRED), the Development Test Command (DTC), and the Natick Soldier Research, Development and Engineering Center (NSRDEC). Phase 4 of the program encompassed work done from January 1, 2006, through February 28, 2007.

6.1.2 Phase 4 Research Objectives

The Honeywell team's Phase 4 program centered on an evaluation conducted with a full platoon of 32 Soldiers at Aberdeen Proving Ground Military Operations in Urban Terrain (MOUT) site in Aberdeen, Maryland. The objective was to assess the cognitive workload classification techniques driven by neurophysiological (EEG) and physiological (ECG) sensors. In a first-ever evaluation of real-time cognitive monitoring in a harsh operational environment, the assessment culminated in a three-phase, 24-hour mission consisting of a coordinated route reconnaissance, a cordon-and-search of a village, and a hasty defense operation. Task load levels were manipulated by introducing unexpected and unplanned events requiring replanning and extensive coordination by the leadership (high task load), as well as lulls in the activity in which partial missions were executed flawlessly with little variation on the preplanned, well versed drill (low task load). Four leaders (platoon leader (PL), platoon sergeant (PSG), squad leader 1 (SL1), and squad leader 2 (SL2)) were equipped with sensors to measure and output cognitive state in real time. The objective of this phase was to test the classification algorithms in a fully operational setting and to explore classification accuracy with EEG, ECG, and a fused EEG and ECG workload classification approach. Overall, the program goal was to demonstrate the viability of real-time cognitive state sensing in a military operational urban terrain environment. The overall objective of the Phase 4 program was to assess Soldier workload levels during various operational tasks requiring different levels of cognitive and physical engagement. The goal was to demonstrate the effectiveness of the AugCog techniques on key leadership positions as measures of cognitive loading during mission phases.

The ACTE was the latest in a series of demonstrations of the Honeywell system of sensors in an outdoor field environment. It advanced the system demonstrated in Phase 2 (see Dorneich et al., 2004) and Phase 3 (see Dorneich, Whitlow, Ververs, Mathan, et al., 2005) by refining the classification algorithms and the experiment design to better assess the classification approach in a true operational environment. The ACTE assessed the effectiveness, specificity, and validity of neurophysiological- and physiological-based measures of cognitive state in an unconstrained, full-mission context utilizing Soldiers as

participants. In addition, Honeywell explored the utility and possible operational feasibility of "closing the loop" via display of the cognitive state information to leaders to allow them to control the flow of information to better match their subordinates' current capacity to process information.

6.2 Phase 4 Challenges

Realizing the vision of an AugCog system in the context of an ambulatory Soldier has been constrained by several challenges. First, as Schmorrow and Kruse (2002) noted, processing and analysis of neurophysiological data have been largely conducted offline by researchers and practitioners. However, for AugCog technologies to work in practical settings, effective and computationally efficient artifact reduction and signal processing solutions are necessary. Second, inferring the cognitive state of users demands pattern recognition solutions that are robust to noise and the inherent nonstationarity in neurophysiological signals (Popivanov & Mineva, 1999). Third, understanding the fluctuations of cognitive state in applied environments requires the development of means to collect reliable neurophysiological data outside the laboratory. Fourth, experiments must be designed, often under conflicting constraints (e.g., operationally realistic tasks vs. well-understood, controlled laboratory tasks), to effectively evaluate classification accuracy. Finally, compact and robust form factors (e.g., size, weight, ruggedness) associated with neurophysiological sensors and processors are a matter of critical concern.

6.2.1 Real-Time Signal Processing Challenges

Conducting military maneuvers in operational environments, such as urban terrain, often does not allow an individual to remain stationary and can demand simultaneous cognitive and physical activity. Consequently, difficulties related to processing of EEG signals in real-world settings include factors associated with both participant motion and the operational environment itself. Thus, utilization of research methods involving EEG in operational environments necessitated the use of real-time algorithms for signal detection and removal of artifacts. Although real-time signal processing and classification of the EEG has been implemented previously (Gevins & Smith, 2003; Berka, Levendowski, Cvetinovic, Petrovic, et al., 2004), it has not been realized in a truly mobile, ambulatory environment.

Inferring cognitive state from noninvasive neurophysiological sensors is a challenging task, even in pristine laboratory environments. High-amplitude artifacts ranging from eye blinks to muscle artifacts and electrical line noise can easily mask the lower amplitude electrical signals associated with cognitive functions. These concerns are particularly pronounced in the context of ongoing efforts to realize neurophysiologically driven adaptive automation for the dismounted ambulatory Soldier. In addition to the typical sources of signal contamination, mobile applications must consider the effects of artifacts induced by shock, cable movement, and gross muscle movement. Specifically, artifacts related to participant motion include high-frequency muscle activity, verbal communication, and ocular artifacts consisting of eye movements and blinks; whereas artifacts related to the operational environment include instrumental artifacts such as electrical noise that create interference with the EEG signal (c.f. Kramer, 1991).

6.2.2 Cognitive State Classification Challenges

The use of EEG as the basis for cognitive state assessment was motivated by characteristics such as good temporal resolution, low invasiveness, low cost, and portability. Although EEG offered several benefits, there were shortcomings related to the noise artifacts described above and the nonstationarity of the neural signal pattern over time. Despite these challenges, research has shown that EEG activity can be used to assess a variety of cognitive states that affect complex task performance. These included working memory (Gevins & Smith, 2000), alertness (Makeig & Jung, 1995), executive control (Garavan, Ross, Li, & Stein, 2000), and visual information processing (Thorpe, Fize, & Marlot, 1996). These findings pointed to the potential for using EEG measurements as the basis for driving adaptive systems that demonstrate a high degree of sensitivity and adaptability to human operators in complex task environments.

6.2.3 Evaluation Challenges

In addition to the practical and system configuration challenges faced when moving from the laboratory to field studies, there were issues of experiment control and the characterization of cognitive state in less constrained environments. It was essential to select tasks that were both operationally relevant and reasonably adaptive to different cognitive task loads. In the laboratory, it was possible to develop simple tasks where workload was manipulated precisely and consistently. Additionally, a user's performance could be collected and evaluated accurately. This made it relatively easy to establish ground truth about a user's likely workload. However, when developing operationally relevant tasks in a field environment, it became substantially harder to manipulate workload precisely and to interpret and assess a user's performance without compromising operational realism. The mobile field evaluation reported herein had two objectives: first, to determine whether an operationally relevant task load manipulation had a measurable impact on a user's workload; second, to establish whether a sensor-based classification approach could effectively classify a user's workload in a harsh operational environment.

6.3 Phase 4 System Design and Architecture

The system constructed to assess the cognitive state classification algorithms consisted of:

- *Sensor hardware:* A variety of sensors collected raw physiological and neurophysiological data, including the ABM EEG system and the Hidalgo Vital Signs Detection System (VSDS) system.
- *Signal processing:* A variety of methods removed artifacts and flag compromised data.
- *Cognitive state classification:* A support vector machine approach assessed cognitive state.
- *Mobile processing:* Mobile, semi-rugged computer platforms processed the raw sensor data into cognitive state classification assessments.
- *Wireless data network:* A wireless data infrastructure sent the classification assessment of subordinates to leaders.

- *Experimenter's base station:* A computing infrastructure and base station controlled the IT component of the experiment and troubleshot any unexpected problems.
- *Mitigation:* A Commander's display allowed human leaders to close the loop.

6.3.1 Sensor Hardware

Each subject was outfitted with an ABM EEG system, a VSDS for cardiac data, a wireless microphone, and a head-tracker.

6.3.1.1 ABM EEG System

EEG data were collected from the ABM EEG sensor headset (Figure 77). The sensor headset acquired six channels of EEG using a bipolar montage. Differential EEG were sampled from bipolar channels CzPOz, FzPOz, F3Cz, F3F4, FzC3, C3C4 at 256 samples per second with a bandpass from 0.5 and 65 Hz (at 3-dB attenuation) obtained digitally with Sigma-Delta A/D converters. Data were transmitted across a Bluetooth RF link to the collection laptop via an RS-232 interface.

Figure 77. ABM's wireless EEG sensor headset.

The sensor headset was developed by ABM as a portable system to record EEG signals. The headset fit snugly on the head and housed EEG sensors like many FDA-approved laboratory EEG systems, such as the Quick-Cap by Neuromedical Supplies or the Electro-Cap by Electro-Cap International. Physiological recordings were made with an experimental eight-channel digital physiological recorder with low-powered EEG and Electro-oculogram (EOG) amplifiers designed specifically for ambulatory recordings. The analog box included input jacks for the electrode leads and event markers, an on/off switch, amplifiers (manufactured and made commercially available by Teledyne Electronics Technologies, Medical Device Group, Marina Del Rey, CA), and optical isolation (designed to meet UL544 requirements). The analog box was coupled to a Real Time Devices microcomputer (commercially available model DSi486SLC, State College, Pennsylvania), which provides A/D conversion, operates the data-acquisition software, and stores the data to a hard drive.

6.3.1.2 *Hidalgo Vital Signs Detection System (VSDS)*

The VSDS (shown in Figure 78) measured heart rate, respiration rate, and body motion and position. The VSDS (Bluetooth-enabled) came with a Bluetooth (Mini Mitter Co. and Hidalgo Ltd.) radio. With the Bluetooth radio, the device was used in full disclosure mode. In this mode, both waveform and summary data were transmitted across a Bluetooth communications link. The document WPSM-IC ATO Phase 2 LSDS Full Disclosure Interface Specification (Howard, 2005) described all data that can be sent by the VSDS. The ACTE utilized the ECG waveform (two views, sampled at 256 Hz) and the three-axis accelerometry waveforms (sampled at 25.6 Hz) signals.

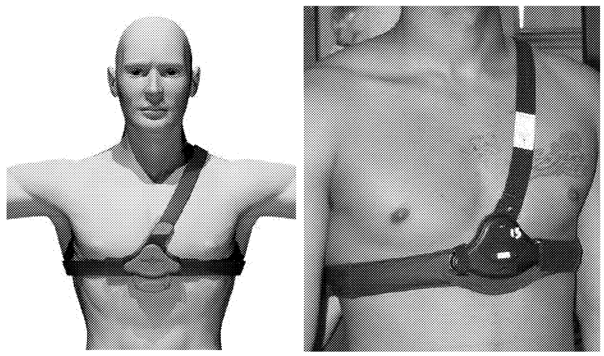


Figure 78. Hidalgo Vital Signs Detection System (VSDS).

6.3.2 Signal Processing

The ABM system supported an independent signal processing stream. Quantification of the EEG in real time was achieved using signal analysis techniques that identified and decontaminated eye blinks and identified and rejected data points contaminated with electromyographic artifacts, amplifier saturation, and/or excursions due to movement artifacts (see Berka, Levendowski, Cvetinovic, Petrovic, et al., 2004, for a detailed description of the artifact decontamination procedures). Decontaminated EEG was then segmented into overlapping 256-data-point windows called overlays. An epoch (the temporal window of analysis) consisted of three consecutive overlays. Fast-Fourier transform (FFT) was applied to each overlay of the decontaminated EEG signal multiplied by the Kaiser window ($\alpha = 6.0$) to compute the power spectral densities (PSDs). The PSD values were adjusted to take into account zero values inserted for artifact-contaminated data points. The PSD between 70 and 128 Hz was used to detect EMG artifact. Overlays with excessive EMG artifacts or with fewer than 128 data points were rejected. The remaining overlays were then averaged to derive PSD for each epoch with a 50% overlapping window. Epochs with two or more overlays with EMG or missing data were classified as invalid. For each channel, PSD values were derived for each 1-Hz bin from 3 to 40 Hz and the total PSD from 3 to 40 Hz. Relative power variables were also computed for each channel and bin using the formula (total band power/total bin power).

6.3.3 Real-Time Cognitive State Classification

Estimates of spectral power formed the input features to a pattern classification system. The classification system used parametric and nonparametric techniques to assess the likely cognitive state on the basis of spectral features, i.e., estimate *p(cognitive state |*

spectral features). The classification process relied on probability density estimates derived from a set of spectral samples. These spectral samples were gathered in conjunction with tasks that were as close as possible to the eventual task environment.

The classification system utilized a support vector machine to discriminate between low and high task load. Support vector machines are linear classifiers that use a quadratic optimization procedure to find an optimal orientation and location for a discriminating hyperplane between two classes. The optimization procedure finds a location and orientation for the hyperplane that lies as far away as possible from examples in each class that are likely to be confused with each other (see Figure 79).

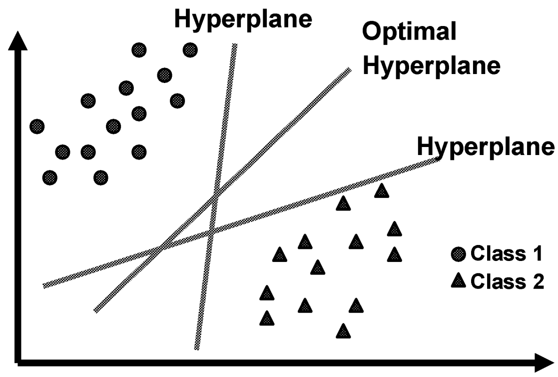


Figure 79. Hyperplane orientation for maximizing generalization (adapted from Takahashi, 2006).

Separating hyperplanes identified using this procedure has been shown to maximize generalization performance (Vapnick, 1999). Although they are linear classifiers, support vector machines were used to solve nonlinear problems by means of the so-called kernel trick. Data that may not have been linearly separable in the original feature space were projected into a high-dimensional space where the data may be linearly separable (Figure 80). The support vector machine used in this effort employed a radial basis function kernel with a kernel parameter of 1 and a slack parameter of 0.05.

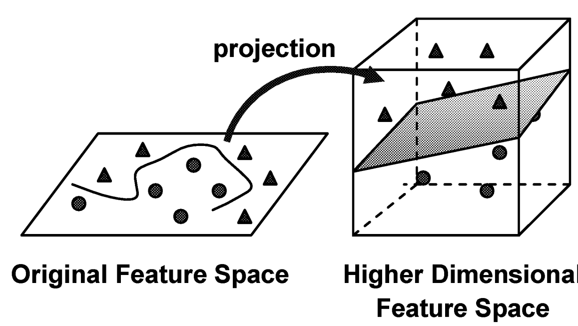


Figure 80. Projection of linearly unseparable data to higher dimensional space in attempt to separate data (adapted from Takahashi, 2006).

6.3.4 Mobile Processing and Data Collection Platform

Each of the four primary Soldier participants (PL, PSG, SL1, and SL2) was followed by a member of the experiment personnel in the role of "shadower." Each shadower remained within the 30 meters of his/her participant to ensure Bluetooth connectivity. Each

shadower carried a specially designed backpack (based on the MOLLE system) that contained a Panasonic Toughbook CF-51 equipped to receive Bluetooth communication from the subject's EEG, ECG, wireless microphone, and head-tracking systems. In addition to logging the data, the raw sensor data were processed on the Toughbook using Honeywell's cognitive state classification algorithms to produce a real-time assessment of the subject's cognitive state. That cognitive state assessment was then transmitted to the base station via the wireless data network (see next section). Additionally, the shadower wore a Web-cam and logged video to the Toughbook. The participant wore a wireless microphone, and the resultant audio stream was multiplexed into the Web-cam video.

The base station, a Toughbook CF-51, received data from the four shadowers' Toughbooks via the wireless data network. The base station was the test team's command and control center of the devices. The base station remotely controlled the four shadower Toughbooks (demonstrating the ability to stop/start processes), monitored processes on four shadower Toughbooks, ran the master radio, remotely troubleshot the shadower Toughbooks, collected data, shut down processes at the end of a trial, and performed other functions.

6.3.5 Wireless Network Connectivity

The ACTE employed a 900-MHz radio modem system to create a wireless data network connecting the four shadowers' Toughbooks to the base station. Sensors on the body were connected via bluetooth to the shadower's backpack, where the heavy processing was done. The resultant information was transmitted wirelessly to the base station. Figure 81 illustrates the connectivity between elements of the system:

- The ABM EEG communicated to the shadower Toughbooks via Bluetooth.
- The Hidalgo VSDS communicated to the shadower Toughbooks via Bluetooth.
- The head-tracker communicated to the shadower Toughbook via Bluetooth.
- The wireless microphone communicated to the shadower Toughbook via Bluetooth.
- The shadower Toughbooks communicated with the base station via 900 MHz data link radios.
- The Commander's Display on the PDA communicated with the base station Toughbook via Bluetooth.

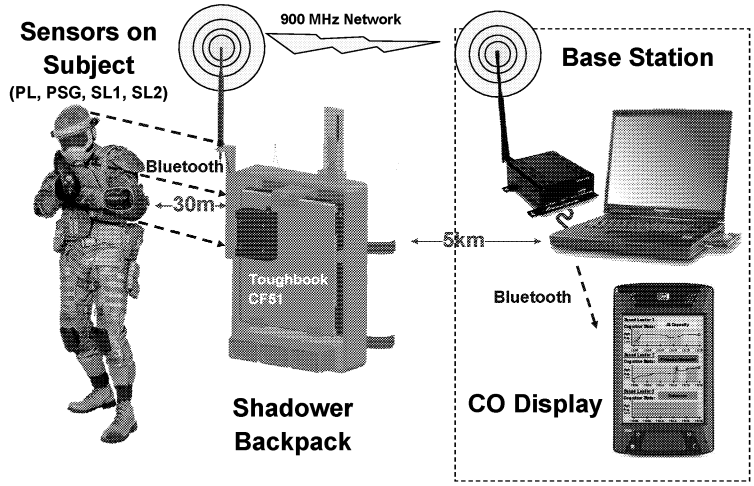


Figure 81. Connectivity between the elements of the wireless data network.

6.3.6 Mitigation Strategies

The objective of the ACTE with regard to mitigations was to explore the utility and possible operational feasibility of "closing the loop" by providing PLs and a company commander (CO) with real-time cognitive state information of subordinate platoon members. This was operationalized by displaying cognitive state information to leaders to allow them to adjust the flow of communications to better match the subordinate's current capacity to process information. In previous evaluations, Honeywell explored using automation to close the loop, where the automation was driven by assessments of cognitive state. In the Phase 4 ACTE, the loop was closed by a human leader using cognitive state feedback of subordinates and then modifying the information flow to those subordinates. This mitigation strategy most closely aligned with the interests of the FFW program, which saw cognitive state feedback as useful information for a leader when assessing the combat readiness of his or her troops.

The ACTE addressed the following questions:

- Would leaders (COs) modify their behavior with subordinates based in part on feedback of the subordinates' cognitive state? If so, how?
- What subordinate cognitive state information was most useful to leaders (e.g., moment-to-moment, trend, etc.)?
- Under what conditions was cognitive state information useful?

In particular, the cognitive states of the PL and the PSG were displayed to the CO during the first part of the 24-hour mission.

Cognitive state information of the subordinates was displayed to the CO via the Commander's Display (see Figure 82). For the ACTE, the Commander's Display relayed information pertaining to the cognitive state of the PL and the PSG. The CO display showed the current real-time assessment of cognitive state via a color-coded text box, where the capacity of the Soldier relative to the task demands was labeled "Unknown" (blue), "Spare Capacity" (green), "At Capacity" (yellow), or "Exceeds Capacity" (red). In

addition, the history of the moment-to-moment assessment of the Soldier's cognitive state was shown via a line graph. The background was redundantly color coded to support "at a glance" processing. The scale of the timeline was user-controllable.

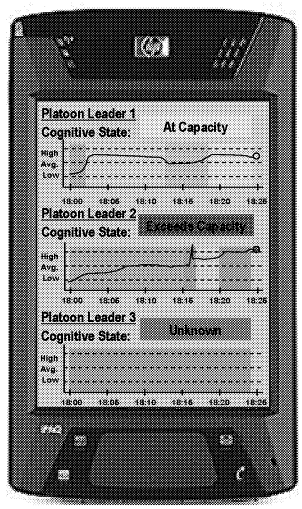


Figure 82. The Commander's Display.

6.3.7 System Integration

Figure 83 illustrates the final data collection system and experiment infrastructure configuration.

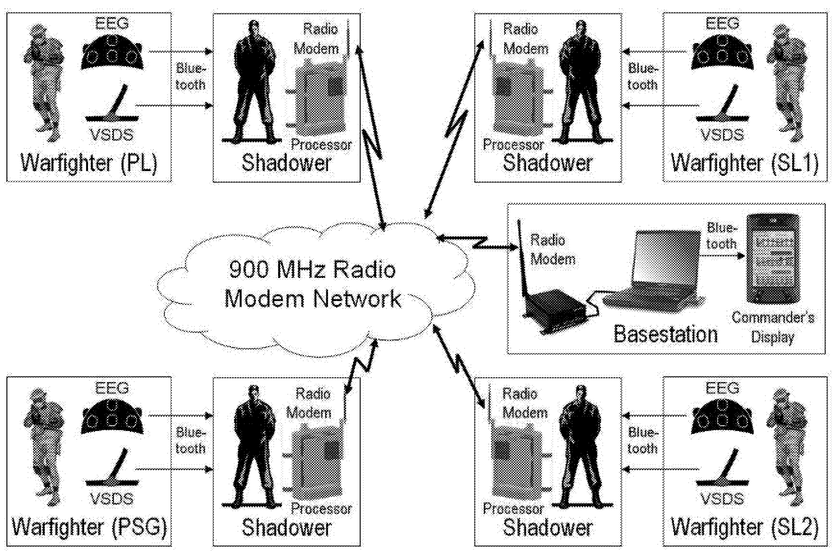


Figure 83. Final data collection system and experiment infrastructure.

Several practical challenges were encountered during the ACTE. First and foremost, the pace of the training was subject to the Soldier's progress through a predefined set of tasks, drills, and procedures. Soldiers were trained to performance on battle drills. The use of simunitions (soap bullets with considerable, but nonlethal, impact velocity) implied that all hardware, including potentially sensitive equipment such as EEG sensors, had to be hardened to withstand a direct hit of a simunition round. In fact, during the experiment, the ABM EEG system sustained a direct hit but was undamaged (see Figure

84a). The weather was another challenge—during two days of training, 12 inches of rain fell (see Figure 84b). The ACTE also required that wireless connectivity be maintained over two networks: the Bluetooth connections between sensors and shadower and the 900-MHz Radio Modem network. Power consumption of the mobile equipment is always a challenge, and battery management was key to ensuring that all devices continued to function despite inevitable delays and schedule changes (see Figure 84c). Finally, EEG sensor integration with the Soldier's standard equipment was a challenge that required special modifications to the padding and padding configuration under the Soldier's helmet (see Figure 84d).

a. b. c. d.

Figure 84. ACTE Challenges: a. Simunitions, b. Weather, c. Power management, d. Sensor integration.

6.4 Phase 4 Augmented Cognition Test Event (ACTE)

6.4.1 Experiment Overview

The ACTE was the next in a series of planned evaluations for aspects of the Honeywell team's ability to assess cognitive state of the mobile Warfighter in an outdoor field environment. The Honeywell effort was concerned with mitigating high-workload demands in the dismounted Soldier environment, especially with regard to information overload due to netted communications. The ACTE evaluated the effectiveness of the classification algorithms to detect the user's cognitive state by correlating classification output to performance in various task load conditions. The ACTE also explored the effectiveness of leaders "closing the loop" via communication pacing to optimally task subordinates.

6.4.2 Operational Scenario

The lead trainer, who also served as the observer/controller (O/C) throughout the evaluation, trained Soldiers on MOUT techniques and battle drills. He conducted two weeks of training, starting with simple entry techniques and progressing to clearing techniques, defensive techniques, and finally battle drills. The Soldiers mastered a technique before moving to the next, as each technique built on what was learned previously. Therefore, the part-mission training tasks were a sequential stepping through of the techniques in Table 37.

Table 37. Simple techniques trained during the part-mission training sessions.

Task #	*ENTRY TECHNIQUES*
1	Ballistic, explosive, and mechanical door breaching techniques
2	Ballistic, explosive, and mechanical window breaching techniques
3	Ballistic and explosive wall breaching techniques
4	Upper-level entry techniques
5	Use of suppression and killing devices to support entry
6	Entry techniques through doors, windows, and walls
Task #	***CLEARING TECHNIQUES***
1	High-intensity versus precision clearing techniques
2	Principles of precision room entry and clearing
3	Principles of precision hall entry and clearing
4	Principles of precision stairwell entry and clearing
5	Reflexive fire techniques
6	Movement techniques within a structure
7	Subterranean considerations
Task #	***DEFENSIVE TECHNIQUES***
1	Hasty defense of an urban area
2	Extended defense of an urban area
3	Defensive considerations (security, protection, dispersion, concealment, fields of fire, covered routes, observation, fire hazards, and depth)

The battle drills, listed in Table 38, were a culmination of all the training that the Soldiers received and allowed the Soldiers to establish their own standard operating procedures. These tasks were not covered until the individual teams and squads demonstrated proficiency in all the basic skills.

Table 38. Battle drills trained during the part-mission training sessions.

Task #	Drill Code	BATTLE DRILLS
1	7-3-D101	Conduct a platoon attack
2	7-3-D108	Enter and clear a building
3	7-3-D112	Conduct initial breach of a mined wire obstacle
4	7-3/4-D103	React to contact
5	7-3/4-D104	Break contact
6	7-3/4-D105	React to ambush
7	7-3/4-D122	React to contact (mounted)
8	7-3-D235	Change formation
9	7-3-D236	Secure at a halt (mounted)
10	7-3-D237	Execute action right or left

There were two principal phases of the 12-day training session during which the Honeywell team collected experiment data. During the period between days 3 and 10, the platoon conducted part-mission training where they repeated a set of tasks for a 3- to 4-hour period. The tasks changed each day. The experiment control ensured that there were definable periods of high and low cognitive workload. Real-time cognitive state classification results were assessed for accuracy during these periods.

The final day of the experiment was a 24-hour, full-mission training session. Again, experiment control ensured that there were multiple periods of definable high and low workload in order to assess cognitive state classification accuracy in these conditions. The 24-hour period was divided into three 8-hour phases:

1. Platoon conducted dismounted movement along the lines of communication to the objective to ensure routes were free of mines and obstacles.

2. On call, Platoon, as part of a larger operation, conducted a cordon-and-search of Objective "Jim" to kill, capture, or expel opposition forces (OPFOR) operating in this urban area, as well as to capture and destroy.

3. Platoon prepared to defend Objective "Jim" for an extended period, and reported any enemy activity in and around this key terrain.

This evaluation focused primarily on the PL, the PSG, and two squad leaders. However, the activities of their subordinates and responses from senior leaders had a direct impact on stress levels experienced by the PL and the PSG.

The platoon-level training exercise used a host of stressors in the MOUT facility. Each is listed in Table 39.

Table 39. Stressors in a MOUT environment.

Category	Example Stressors
Loss of sight	Distributed squads
Confusion	Changes in the plans, conditions, and mission; loss of communications
Realism	Extended operational period (e.g., 24 hours of operation) in the urban facility
Fatigue	Extended movement to the facility followed by an assault and then occupation of the site for long periods in a defensive posture
Uncertainty/Threat	Use of OPFOR to prevent friendly BLUFOR from gaining control of the urban facility to “hit” the BLUFOR at different times
Evaluation Stress	Use of simunitions
Surprise	Imposition of unexpected elements that affect plan
Severe Weather	Periods of high heat and humidity; intense rainfall

6.4.3 Experiment Objectives

The fully equipped Soldier/participant was outfitted with a mobile sensor-based ensemble that monitored his/her cognitive and attentional state. Experimental tasks placed participants in conditions of high and low workload by manipulating task load. Over the course of the training run, the classification system developed a model of power spectrum profiles associated with various cognitive states of interest. During experiment runs, the classifier examined each power spectrum estimate and associated it with the most likely cognitive state. The output of the model was an assessment of a participant's cognitive load. The classification analysis focused on the following questions:

- *Bias, variance, and temporal smoothing:*
    - How well did the classifier fit and discriminate between workload classes in an inherently noisy and dynamic environment?
    - How well did the classifier generalize to unseen data over spans of tens of minutes—when task characteristics remained the same?
    - Did classification accuracy improve as the output of the classifier was integrated over time?
- *Discriminating features:* What aspects of EEG signal served to discriminate between high and low workload?
- *Fusion:* Was overall classification accuracy improved by integrating additional sensor sources?
- *Sensor density:* How many channels of EEG signals were required for accurate classification?
- *Long-term generalization:* How well was the classifier likely to generalize over time spans of days as the task context and patterns of general physiological activity changed (sleep, stimulants), etc.?

In addition, Honeywell explored the utility and possible operational feasibility of “closing the loop” via display of the cognitive state information to leaders to allow them to control the flow of communications to better match their subordinates’ current capacity to process information. The research addressed the following questions:

- Did leaders (CO) modify their behavior with subordinates based in part on feedback of the subordinate's cognitive state? If so, how?
- What subordinate cognitive state information was most useful to leaders (e.g., moment-to-moment, trend, etc.)?
- Under what conditions was cognitive state information useful?

6.4.4 Experiment Hypothesis

The objectives of the ACTE were threefold. First, could reliable EEG and ECG measures be taken in the field under mobile, combat conditions? Second, if reliable signals were collected, could Honeywell's cognitive state classification algorithms provide meaningful assessment of cognitive state? Third, how would commanders use cognitive state feedback of subordinates to mitigate their subordinates' workload and optimize information flow?

Experimentally, the principal hypothesis tested in the ACTE was as follows:

> *The Honeywell cognitive state classification algorithms would be able to differentiate periods of high and low cognitive workload using a combination of physiological (ECG) and neurophysiological (EEG) sensors.*

6.4.5 Experiment Design

The independent variable in the ACTE was workload (all phases). The experiment scenarios were manipulated to ensure definable periods of high and low cognitive workload. Periods of low workload included completing initial paperwork, reporting activities, preplanning, conducting long hasty defenses, consolidation/transition, after action reviews, and periods of low activity during missions. High-workload periods were characterized by multiple task performance under time pressure and fatigue. Examples of high workload were replanning due to change in circumstances (e.g., enemy location, available squads, loss of communication, etc.), directing squad movements during pre-assault, squads in assault, managing multiple communications (i.e., responding to commanders, squad leaders, or other PLs), or calls for fire/backup. Stressors that contributed to high workload included a degree of frustration or stress, loss of communication, lack of asset availability, and loss of situation awareness (SA) of squad locations and activities.

6.4.6 Dependent Measures

The objective of this experiment was to assess the ability to classify cognitive-state-based EEG and ECG sensor data. The principal issue in the experiment design was to create detectable and sustained (5-10 minutes) high or low workload multiple times within any single training session. Definable periods of high and low workload, known here as the "ground truth" of actual workload sustained by the participants, was determined by a team of experts based on task breakdowns, experimenters' observations, video review, and post-session Soldier interviews. The output of the cognitive state classification algorithms was compared with the ground truth workload to determine classification accuracy.

6.4.7 Participants

The ACTE utilized a full platoon of Soldiers from the North Carolina National Guard (NCNG) Combined Arms Battalion, as shown in Figure 85.

Data were collected from four participants: the PL, the PSG, the squad 1 leader (SL1), and the squad 2 leader (SL2). Each of these four participants were outfitted with an ABM EEG system, a Hidalgo VSDS ECG system, a wireless microphone, and a head-tracker (not shown). All other Soldiers were outfitted with the Hidalgo VSDS system as part of a coordinated, parallel experiment led by USARIEM. Three squads participated, each with approximately nine Soldiers. OPFOR were staffed by remaining members of the NCNG. Onsite ATC Soldiers served as part of the test team. The NCNG company commander was also a member of the test team.

At the platoon level, participants in the MOUT training ranged in age from 21 to 40 (average age = 27.2 years), with army experience ranging from 0.5 to 18.3 years (average = 7.6 years). All Soldiers were male. Of the 32 Soldiers in the platoon, 28 had seen combat. None of the Soldiers had previously trained at the Aberdeen Proving Ground.

Four participants wore AugCog sensors. The PL had 15 years of Army experience, although he was new to the PL position. The PSG had 16 years of Army experience, the SL1 had 17.5 years of Army experience, and the SL2 had 8.9 years of Army experience. The average age of the four AugCog participants was 33.2 years (ranging from 25 to 40). All had seen combat.

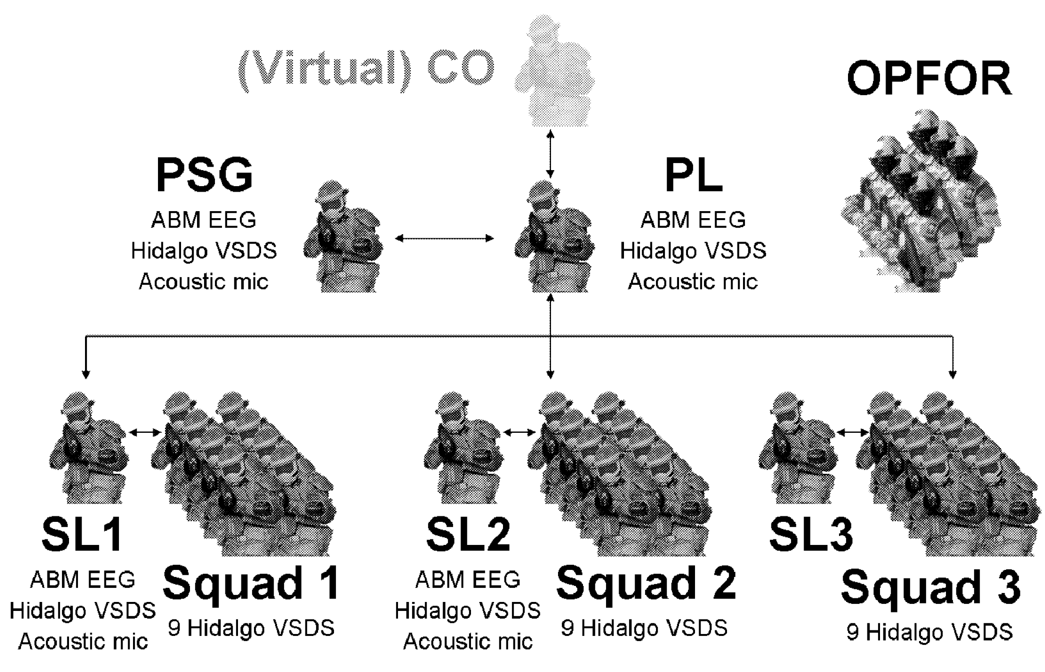


Figure 85. Platoon participants and the equipment they wore.

6.4.8 Experiment Protocol

The ATC is a Major Range and Test Facility Base (MRTFB), operating under the guidance of the Department of Defense (DoD). The ATC's primary mission is to support DoD test and evaluation requirements. The ATC also conducts testing for federal, state,

and local governments, academia, private industry, and foreign governments (U.S. Army Aberdeen Test Center, 2007).

The ATC MOUT training facility, known as Mulberry Point, contains a pre-assault staging and assault areas. It is a compound with several single- and multi-story buildings, with windows, doors, and hallways. The site serves as a close-area combat training ground. The test site is equipped for data collection, including cameras in and around buildings.

6.4.9 Experiment Schedule

Table 40 contains the experiment schedule of the ACTE.

Table 40. ACTE experiment schedule.

Day	Day	AM (0:00-12:00)	PM (12:00-24:00)	Note
1	Monday	Travel day		
2	Tuesday	Paperwork and questionnaire	AugCog system check, possibly data collection	
3	Wednesday	AugCog data collection 3-4 hrs		Part-mission
4	Thursday	AugCog data collection 3-4 hrs		Part-mission
5	Friday	AugCog data collection 3-4 hrs		Part-mission
6	Saturday	Off day		
7	Sunday	Off day		
8	Monday	AugCog data collection 3-4 hrs		Part-mission
9	Tuesday	AugCog data collection 3-4 hrs	Soldiers sent home to sleep	Part-mission
10	Wednesday	0:00 start of 24-hour freeplay	24:00 end 24-hour freeplay	Full-Mission
11	Thursday	Everyone sleeps	AugCog wrap up	
12	Friday	Travel day		

6.4.10 Accuracy Metric Methodology

The metric used to evaluate classification performance is the area under the receiver operating characteristic (ROC) curve (see Duda, Stork, & Hart, 2001). ROC curves plot true positives (on the y-axis) against false positives (on the x-axis) as a threshold for discriminating between targets and distracters. The ROC curve provides a way to assess the degree of overlap between two univariate distributions. It is widely used to evaluate human and machine signal-detection capabilities. The ROC curve provides a way to assess the degree of overlap between the output of a classifier for two classes of data. Perfect classification produces an area under the curve value (Az) of 1.0, whereas chance performance produces an Az value of 0.5.

6.4.10.1 Ground Truth Assessment

To calculate the accuracy of the classification approach, classifier results are compared with ground truth. *Ground truth* is defined as the actual workload experienced by the participant at any given moment. The output of the classifier at any moment is then compared with the ground truth to determine the accuracy of the classifier, as described in Section 6.4.10.

In the operational setting of the ACTE, it was not possible to vary the workload directly. Instead, varying degrees of task load induced varying amounts of cognitive workload. Furthermore, the amount of cognitive workload induced in a participant is a function not only of the task load, but also of factors such as stress, fatigue, training, experience, and individual differences in capabilities. Thus, there is no way to directly correlate task load to workload in a systematic way to derive ground truth.

During the ACTE, multiple streams of data were collected with the objective of providing experts enough insight to make a determination of ground truth levels of workload for each participant in each scenario. Data included:

1. Video from a roaming camcorder, focused on the platoon-level action
2. Video from the web-cam of the shadower, focused on the participants
3. Notes from an observer at a central (video) monitoring site
4. Annotations radioed in from the shadower and entered at the base station into the time-stamped data stream via an Annotator's graphical user interface (GUI)
5. Post-scenario cognitive walk-through with the participants as they reviewed the video of the day's events with an experimenter
6. Post-scenario NASA TLX (Task Load Index) surveys and questionnaires

Not all data were collected for every part-mission and full-mission scenario, but some combination of data streams was available for expert review. The notes, annotations, and cognitive walk-through feedback data streams were merged (by time-stamp) into a spreadsheet. Two experts then independently reviewed the various video streams, taking into account various other data sources, to make a moment-to-moment assessment of the cognitive workload experienced by the participant at any given time-stamp. The result was a time-stamped series of blocks of low, medium, or high cognitive workload. Physical load was also assessed by the experts. Their respective results were then compared with gain a measure of inter-rater reliability on the cognitive workload assessments of ground truth.

Operationally, low workload was defined to be times when the participant would have been able to take on additional cognitive tasks. Medium workload was determined to be times when the participant was cognitively engaged but able to handle the task demands. Finally, high workload was defined as times when the participant was unable to take on any additional tasks and, in fact, was unable to handle the current task load to the best of his or her ability.

A final canonical assessment of ground truth was created by reconciling the two individual experts' assessments. Time periods of disagreement were flagged. The two experts then jointly reviewed the video and other data streams to make a final assessment of the workload in the disputed block. When no consensus could be reached, a third rater was brought in to resolve the disagreement; however, this option was never needed. The reconciled ground truth tables were used to calculate the accuracy metric of the classification algorithms.

6.5 *Phase 4 ACTE Results*

6.5.1 Training Effectiveness

6.5.1.1 Part-Mission Scenarios

The Soldiers were given questionnaires before and after the training. Figure 86 illustrates the improvements in ratings before and after training, based on the Soldiers' subjective ratings, over a range of MOUT tasks.

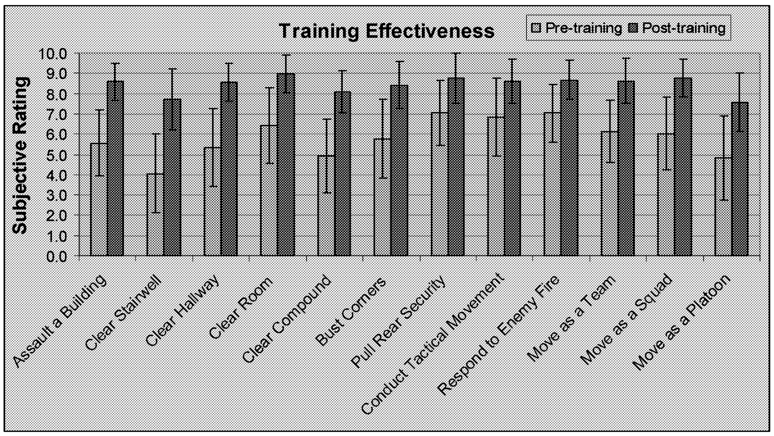


Figure 86. Subjective ratings of training effectiveness (bars represent standard deviation).

6.5.1.2 Full-Mission Scenarios

The final day of the training exercises was a 24-hour full-mission training session. Soldiers used the techniques and skills learned during the part-mission training sessions. The Soldiers reported their level of fatigue before and after the 24-hour mission. They reported a fatigue of 3.7 (standard deviation 2.3) before the mission, and a fatigue of 6.8 (standard deviation = 2.0) after the mission. They were asked about their mission effectiveness after the 24-hour mission. Figure 87 reports the subjective ratings of tasks by the Soldiers.

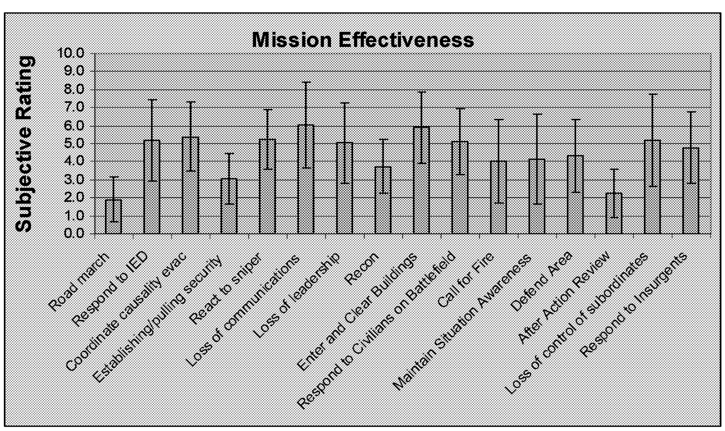


Figure 87. Mission effectiveness after the full-mission scenario.

6.5.2 Ground Truth Inter-Rater Agreement

Two experts independently performed the ground truth analysis described earlier. Their respective results were then compared with gain a measure of inter-rater reliability on the cognitive workload assessments of ground truth. For the data sets analyzed, agreement between the raters was high. Agreement in the rating of physical load was 94.9%. Agreement in the rating of cognitive workload was 87.9%.

6.5.3 Cognitive State Classification Results

The classification analysis described herein focused on data from a part-mission scenario run the day before the full-day mission. The training was a full-platoon, force-on-force (i.e., OPFOR with simunitions) exercise that involved the full gamut of a mission. The data set included waiting, moving out, approaching and coordinating an attack of the compound, clearing buildings A and B and the inner courtyard, holding the compound, and coordinating medical evacuations. SL2 was killed during the scenario, and the leadership needed to adjust. The data from this day were chosen for analysis because:

- A full and complete set of both cardiac and EEG data were available for more than one participant,
- There were distinct and prolonged periods of high and low workload, and
- Physical activity in both low- and high-workload conditions was similar—reducing the potential for confounds associated with physical activity.

6.5.3.1 Bias, Variance, and Temporal Smoothing

A major concern in the environments in which dismounted Soldiers function is that noise from myriad sources could completely mask features that could be used to discriminate between high and low workload. Thus, a classifier may fail to adequately discriminate between workload classes. The capacity of a classifier to accurately fit the training data is known as the *bias* of the classifier. There is also concern that these noise characteristics could change dramatically over time—so that even if a classifier is able to effectively discriminate between workload classes over a short temporal window, it fails to adequately generalize to unseen data collected a few seconds or minutes beyond the duration of the data used to train the classifier. The capacity of a classifier to generalize is referred to as the *variance* of the classifier.

One way to explore the bias and variance of a classifier was through a process called n-fold cross-validation. This procedure entailed splitting the data into N subsets. At each iteration of the validation procedure, one of these subsets (N_i) was used for testing the classifier, while the remaining 1 – 1/N sets were used for training the classifier. A typical choice of N was ten. Estimates of bias and variance get more conservative as the size of n decreases. The classifier had to be trained with less of the data and was assessed by generalizing to a larger subset of unseen data. The AugCog team assessed its classification approach with two individuals: the PL and the PSG—using both the widely used ten-fold cross-validation approach and the more conservative two-fold cross-validation procedure.

In noisy operational environments, EEG and other electrophysiological sensors could be compromised by noise over short temporal windows. One strategy for dealing with momentary fluctuations in classification accuracy was to median filter the output of the classifier over different time windows. One consequence of temporal smoothing of classifier output was to introduce a lag in the decision process. The analysis considered the tradeoff in accuracy as the temporal window of output smoothing was varied.

As Figure 88 (left) illustrates, base EEG classification accuracy for the PL ranged from 0.76 (using two-fold cross-validation) to 0.83 (using ten-fold cross-validation). Base results for the PSG ranged from 0.66 (using two-fold cross-validation) to 0.75 (using ten-fold cross-validation), as seen in Figure 88 (right). Accuracy for both Soldiers rose monotonically up to a one-minute-long temporal smoothing window. However, the rate at which temporal smoothing benefited accuracy diminished beyond approximately two to three seconds of smoothing.

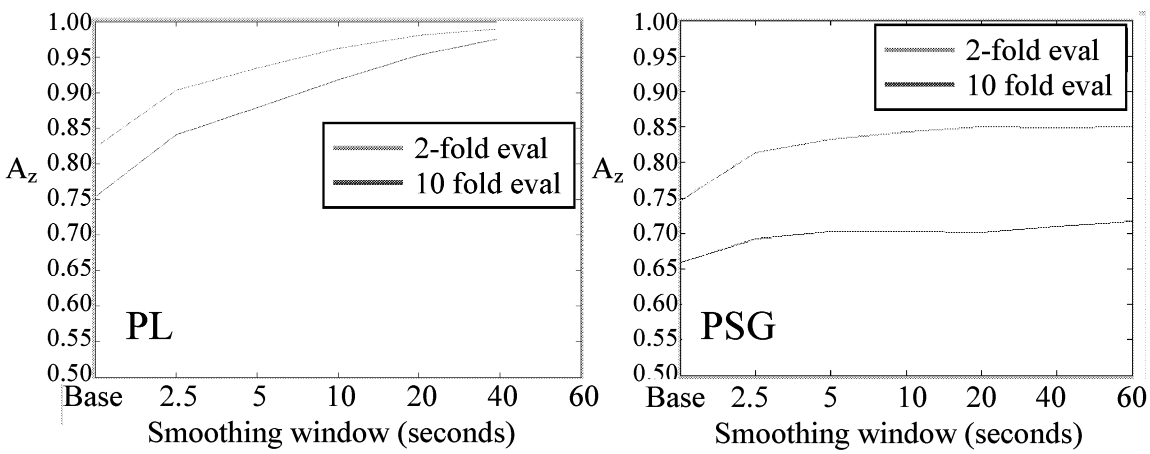


Figure 88. EEG-based classification accuracy for the PL (left) and the PSG (right) as a function of validation technique and temporal smoothing window.

The discrepancy between the more conservative two-fold cross-validation and the more optimistic ten-fold cross-validation was more pronounced for the PSG than it was for the PL. This could have indicated some change in the features that served to discriminate between high and low workload over time; these changes could have stemmed from changes in task, strategy, artifacts, or a variety of physiological factors.

6.5.3.2 Discriminating Features

The analysis also included a qualitative examination of the spectral features that served to discriminate between high and low workload. Figure 89 depicts the PSD estimates for high and low workload across six channels of EEG for the PL (Figure 89 left) and the PSG (Figure 89 right). Each graph in each of the figures represents a channel. The x-axis in each graph represents frequency; whereas the y-axis represents amplitude. The red line in each graph represents averaged spectral power in the high-workload condition; whereas the green line represents average spectral power in the low-workload condition. The blue line in each graph corresponds to the mean spectral power across both high- and low-workload conditions.

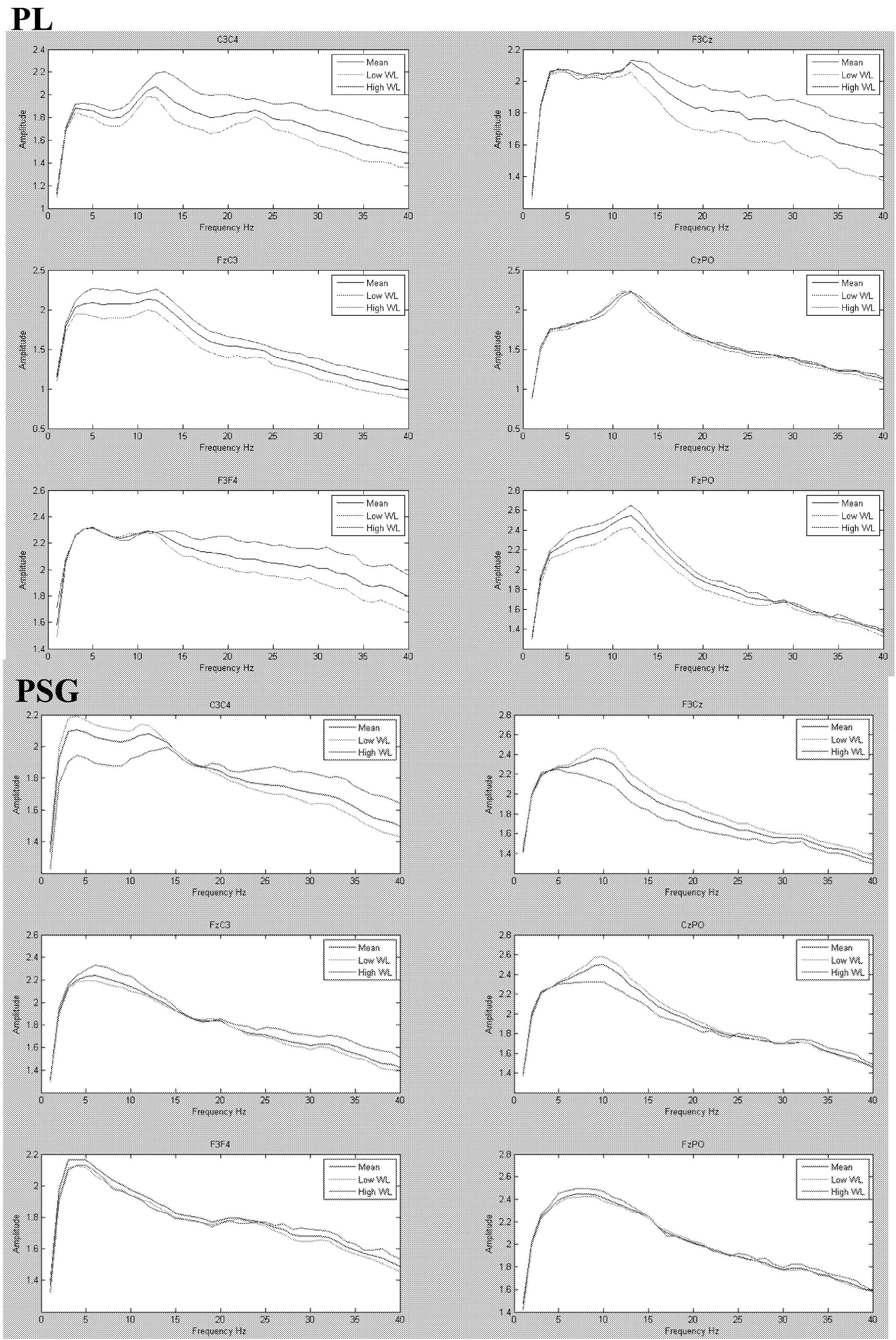


Figure 89. PSDs in each band for the PL (upper) and PSG (lower).

An analysis of graphs for both participants suggested that power in the beta (12 to 30 Hz) and gamma (30 to 40 Hz) bands was the most discriminative feature for both subjects. However, this pattern was most pronounced for the PL and may have accounted for the superior classification results observed relative to the PSG. This discrepancy across individuals also points to the importance of an individualized approach to classification, instead of an approach that relies on group norms.

6.5.3.3 Sensor Fusion

One strategy for robust classification in noisy field environments is to fuse data from multiple sources. Such an approach exploits the joint strengths of different data sources while minimizing their individual weaknesses. One approach for integrating multiple sensor sources is to integrate information from multiple sensor sources into a common feature vector and allow a classifier to find an optimal weighing for each feature based on the training data.

Honeywell assessed the effect of including the interbeat interval (IBI) estimates as a feature for classification using a cross-validation procedure. The fusion of cardiac data provided a substantive boost to overall classification performance. These improvements were most pronounced for the PSG, as seen in Figure 90. Base classification for the PL went up from 0.76 (using two-fold cross-validation) to 0.83 (using ten-fold cross-validation) to 0.87 (two-fold) and 0.95 (ten-fold). Base classification for the PSG went up from 0.66 (using two-fold cross-validation) to 0.75 (using ten-fold cross-validation) to 0.83 (two-fold) and 0.86 (ten-fold), respectively.

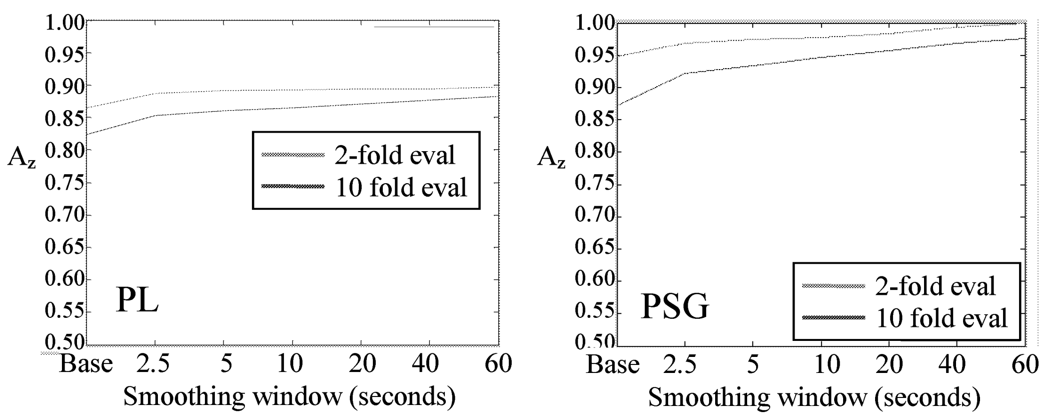


Figure 90. Classification accuracy for the fused sensor data for the PL (left) and the PSG (right).

6.5.3.4 Sensor Density

The EEG system used in the field evaluation consisted of a six-channel system. Spectrally decomposing the data from each channel yielded a 30-feature vector. Several potential problems are associated with working with data of such a high dimensionality. These problems and potential remedies for them include:

- *Limited insight into phenomena*: High-dimensional data can obscure the underlying psychophysiological phenomena from researchers. Classification outcomes provided little information about the specific features that help discriminate between targets and distracters on the basis of EEG. Identifying a

subset of features that help discriminate among conditions could provide insight into cognitive processes associated with perceptual judgments.

- *Potential for poor generalization*: Identifying a smaller subset of features could contribute to better generalization performance. By basing classification on a subset of features that discriminate between classes, it may be possible to reduce the possibility of poor generalization performance as a result of spurious activity along irrelevant dimensions.
- *Computational inefficiency*: Working with an optimal subset of data could improve real-time performance of the classification system and reduce computer processing, memory, and data storage requirements.
- *Limited user acceptance*: Identifying a subset of informative channels could lead to EEG systems that are less cumbersome to configure and consequently more acceptable to users of an EEG-based triage platform. The fewer the number of channels that are necessary for effective performance, the less time required for setup and the better the comfort for users.

6.5.3.5 Dimensionality Reduction Approaches

Feature selection methods fell into two broad classes: filter methods and wrapper methods.

Filter methods: Filter methods select a subset of features based on the intrinsic properties of data. Principal component analysis is an example of such a method. Although filter methods can often be computationally efficient, there is little guarantee that such an approach will produce feature subsets that improve discrimination between classes. Filter methods solve a different problem from the one the classifier will be solving.

Wrapper methods: Wrapper approaches explore the efficacy of different combinations of feature subsets in solving the required classification problem. Although exhaustive search of every combination of features would provide the best answer, it is rarely a practical option because the resulting search process is exponentially long (2^d, where d represents the dimensionality of the data). However, heuristic search procedures such as backward and forward elimination generally provide good solutions. Under backward elimination, one starts with all the features present and, at each iteration, eliminates the feature whose exclusion produces the best validation performance. Under forward elimination, one starts with a single feature and, at each iteration, adds a feature whose inclusion provides the best classification performance. Backward elimination may provide better solutions with patterns where interactions between dimensions may be critical to making effective discriminations between classes—particularly in cases where a feature's individual contribution may be weak.

6.5.3.6 EEG Channel Selection with Backward Elimination

The focus of this analysis was to identify a subset of EEG channels using backward elimination. The objective of this analysis was to find a subset of channels that could match or exceed the performance of all channels together. With each iteration of the ranking algorithm, each channel of the current channel set was sequentially eliminated from consideration. The channel whose exclusion led to the best performance results was

eliminated from further consideration. The ranking assigned to each channel corresponded to the order in which it was eliminated. The first channel to be eliminated was ranked as being last in importance; whereas the last channel to remain in consideration was regarded as being of the highest importance. The performance of each feature subset was assessed using ten-fold cross-validation. The performance metric used was the area under the receiver operating curve (Az). The channel ranking procedure produced channel ranks for each subject. Figure 91 (PL left and PSG right) plots classification accuracy as a function of the top n channels.

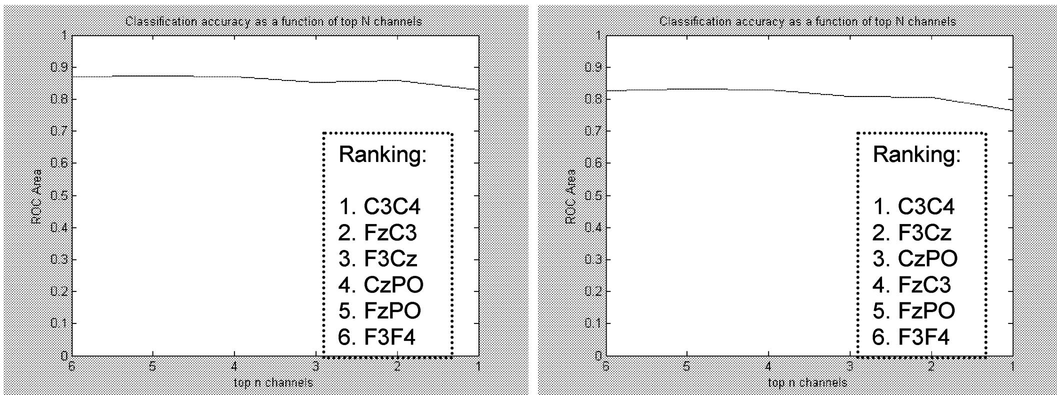


Figure 91. Classification accuracy as a function of the top n channels.

6.5.3.7 Most Salient Channels

The channel ranking procedure yielded a consistent set of features for both subjects. Classification performance suffered little with the exclusion of all but the two most salient channels. These top channels were identical for both participants (C3C4). This channel was located right at the apex of the skull and is likely to have been least affected by helmet-related artifacts because of good clearance between the sensors and the helmet at these locations.

Although these results require further validation, they suggest that accurate workload classification may be feasible with as few as one or two sensors. This has compelling implications for the design of practical EEG systems that could be easily integrated within helmets and could generate broader user acceptance.

6.5.3.8 Long-Term Generalization

Although the results presented above suggested that robust and accurate classification was feasible in the field, a qualitative analysis of longitudinal data spanning days suggested that much more research is necessary to create classifiers that can generalize over time spans of days as the task context and patterns of general physiological activity change. For example, Figure 92 (upper) contrasts spectral data associated with entering and clearing a building during a morning session after a full night of sleep with spectral data in a task where high workload was induced by communications load following a night of sleep and food deprivation (Figure 92 lower). The graphs in Figure 92 show dramatic differences in the spectral profile associated with high and low workload across the two days.

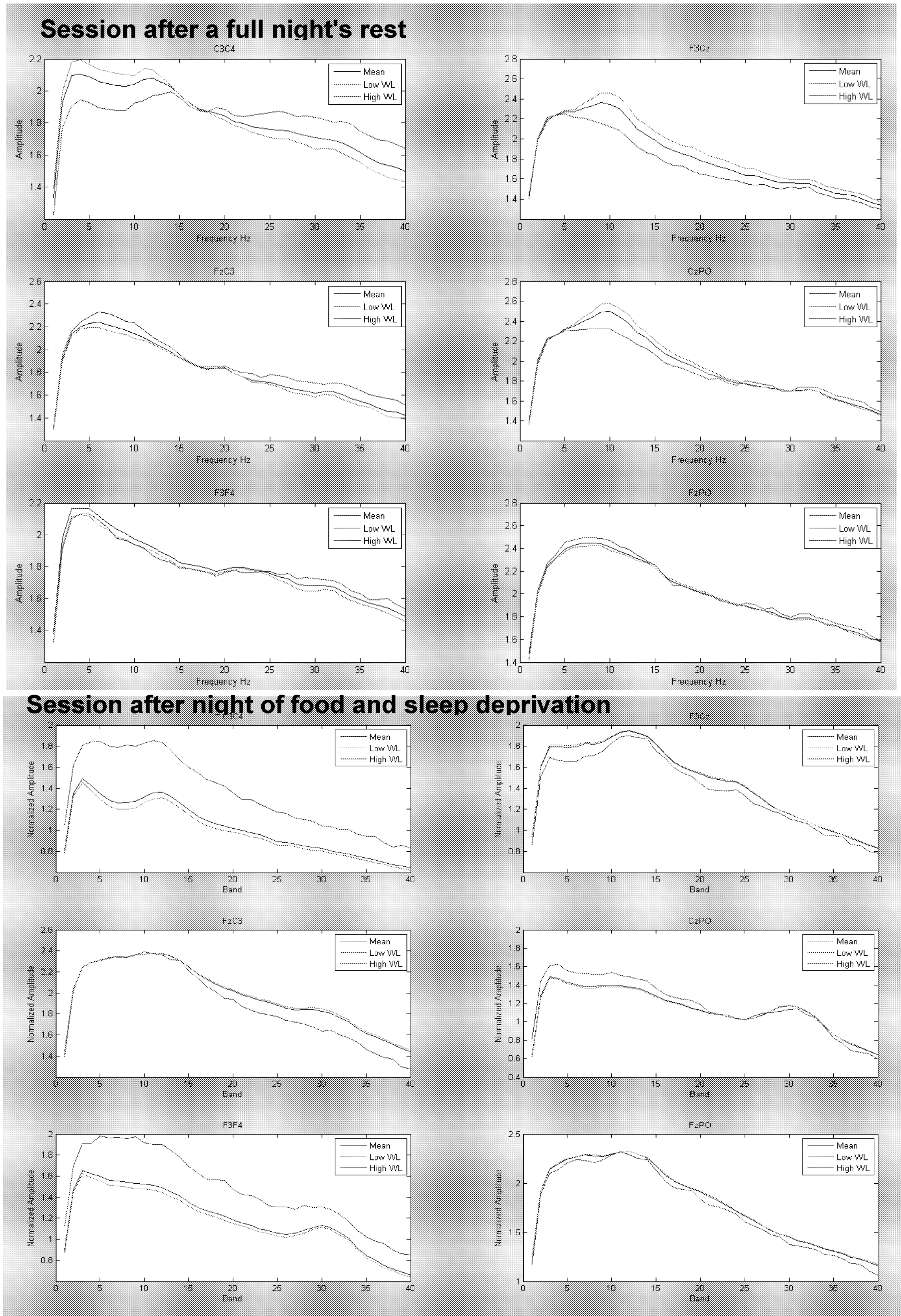


Figure 92. Spectral data after full night's rest (upper) and after night of sleep/food deprivation (lower).

6.5.4 Commander's Display Feedback

The ACTE collected subjective feedback from the CO after he was presented with the Commander's Display during the first and second phases of the full-mission scenario. In general, he found the Commander's Display realistic and useful, especially after the PSG was eliminated due to sniper fire. He primarily used the current status of cognitive state, and found the history graph of limited usefulness, as reflected in his ratings shown in Figure 93. He particularly felt that it allowed him to understand what was transpiring in the field, especially when communications broke down.

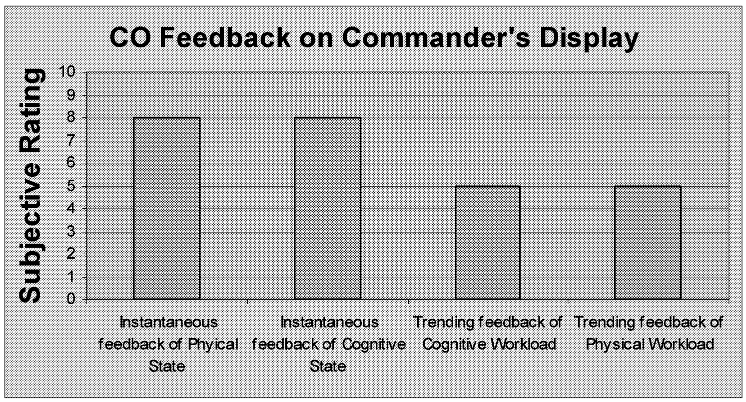


Figure 93. CO subjective ratings of the Commander's Display.

Figure 94 illustrates the ratings of usefulness by task, on a scale of 1 to 10 where 1 is "not useful" and 10 is "very useful." He felt that the physical workload description was particularly useful during "React to contact (IED)" and reiterated that this type of feedback is "good during any time comms go down--Very, very useful."

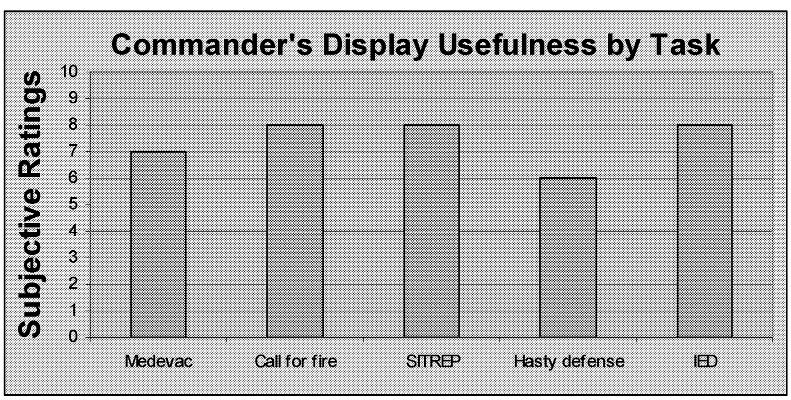


Figure 94. CO subjective ratings of the usefulness of Commander's Display by task.

The PL was also given a ruggedized Commander's Display to take into the field during the ACTE. Unfortunately, connectivity issues precluded its use. However, he offered his initial reaction and opinion that this sort of display is not really meant for his level. He felt the PL's job was to "shoot, move, communicate." He felt it might be more appropriate for a medic to see who is down (i.e., incapacitated) and who talks to the battalion. If more than one was available, he felt the Company Commander or Company Executive Officer in the tactical operations center should have one. At the company level,

there should be one in the command net with the administration and logistics net. At the platoon level, if anyone should have a commander's display, it should be the PSG.

6.6 Phase 4 Discussion

6.6.1 Transition to the Army

The Honeywell AugCog team has been working with the U.S. Army since the inception of the Defense Advanced Research Project Agency (DARPA) Improving Warfighter Information Intake Under Stress (IWIIUS)/AugCog Phase II program. As Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) capabilities enable unparalleled information sharing and real-time collaboration across geographically diverse assets, the concern is the impact on the individual Soldier. When deployed correctly, the technologies will provide greater situational understanding for decisive actions; however, the success will be dependent on the Warfighter's ability to sort through the vast array of continuous information afforded by a full range of netted communications. The Army recognizes the potential strain the added capabilities will impose on deployed Soldiers operating in the stressful conditions of war. Therefore, as new systems are spun into the Army's Ground Soldier System (GSS) program, requirements exist for systems to be developed to assist Soldiers during all operational conditions, particularly when the Soldiers' cognitive skills are degraded, such as during sleep deprivation. The first step is recognizing when these degraded cognitive states exist. AugCog technologies offer the ability to detect degraded performance states.

The Army is also keenly aware that advances in technology often impose additional cognitive, physical, or decision-making requirements on the Soldier. The Future Combat System (FCS) program requires specific cognitive engineering analyses for the new systems developed for the Soldiers. Through Soldier testing and/or modeling, the new systems must demonstrate the capability to minimize physical and cognitive workload and to establish performance standards. The advanced technologies developed and tested throughout the Honeywell AugCog program can be tailored as evaluation tools to assess the cognitive load imposed by new system designs. In the same way that new technologies will be tested, the enhanced roles and responsibilities brought on by the additional capabilities will also need to be thoroughly evaluated for their effects on cognitive processing required. Honeywell, in an effort to further demonstrate the efficacy of the AugCog cognitive classification techniques as a tool for workload assessment, demonstrated the AugCog system at the C4ISR On-The-Move in the summer of 2007, as part of FFW Advanced Technology Demonstration (ATD). In future technology efforts, Honeywell will equip leaders in a platoon to demonstrate real-time cognitive state assessment while the Soldiers are demonstrating the latest technological advances in warfighting.

6.6.2 Physiological- and Neurophysiological-Based Classification

This latest Honeywell evaluation demonstrated not only the additional benefits afforded by using the fused physiological- and neurophysiological-based classification, but also the potential for basing cognitive state assessment on physiological measures (i.e., IBI) alone. Through enhanced signal processing to correct for artifacts in the cardiac signal, the use of the VSDS output for cognitive state monitoring is quite promising. These

sensors are currently being tested as part of FFW's Physiological Status Monitor (PSM) program in the FFW Increment 2 upgrade. Therefore, cognitive state monitoring techniques developed as part of the AugCog program have the potential to be deployed in systems being offered to Soldiers in just a few years' time.

Neurophysiological sensing such as that enabled by EEG sensors is still a few years away. The Army is investigating the development of new sensors, such as the dry electrode, that will enable EEG to be fieldable in future upgrades to Army system ensembles. Follow-on work is being funded by the U.S. Army NSRDEC to begin further development with a goal of transitioning the technology to the Army. Adding to the fieldability were the promising results provided by the current evaluation that demonstrated high degrees of cognitive state classification accuracy with a minimal set of electrodes. Findings such as these further the feasibility of deploying AugCog technologies in the near future.

7 Program Wrap-up

7.1 Evolution of a Mobile Classification Ensemble

Efficiency advances in signal processing and classification techniques, the paring down to the most effective and practical physiologically-based sensing technologies, and the miniaturization of the sensing components have led to a remarkable transformation from the laboratory-based system to the current mobile classification ensemble. Developments in dry electrodes and helmet integration will further improve the capability to deploy these systems in operational environments. The next engineering developments for the Honeywell AugCog program will be to integrate the capability and classification outputs into the network-centric information environment afforded by future military operations. As part of the C4ISR On-The-Move (OTM) demonstration, the Honeywell program will be working to further improve the processing efficiency to the point where the processing, which was previously hosted on the Toughbook carried by the shadower in the Augmented Cognitive Test Event (ACTE), can be ported to an on-the-body processor. The cognitive state classification needs to be power- and processor- aware so that it will not unnecessarily drain key processing capabilities but will provide enough capability to ensure real-time cognitive state monitoring. The wireless cognitive state classification output will be made available by the OTM Future Force Warrior (FFW) leader systems, enabling the broadcast over the entire communications network. The cognitive state would be tied to the individual Soldier as a node in the larger system.

Additional work to further enhance the situational understanding of the individual Soldier will be to couple the cognitive state information with context-aware sensors to truly gain the total picture. Context gathered from such sensors as accelerometers indicating body position and/or rifle position will further inform whether the Soldier's current cognitive state is appropriately matched to the situation.

Phase 4 culminated with the demonstration of a mobile AugCog system in an operational context, thus taking the first step toward a full evaluation in a realistic Army operational setting. Several challenges must be met to take this next step. The remainder of this section outlines these challenges.

7.2 System Deployment Challenges

As the Honeywell Augmented Cognition (AugCog) team transitions from Honeywell's mobile, experiment scenarios to future battle lab integration events, it will begin tailoring the Honeywell AugCog system of systems to address likely deployment challenges. Feedback from Honeywell's Army partners indicates that the Honeywell sensor and computational component must address the following high-level requirements:

- Provide reliable performance under harsh dismounted conditions
- Integrate with Army subsystems with no appreciable increase in weight, size, power consumption, network bandwidth utilization, or computational resources
- Garner very high levels of user acceptance and operational acceptance

7.2.1 System Reliability

Maintaining system reliability under harsh conditions is the reality of the dismounted Soldier domain. In addition to the common challenge for all electronics in the battlefield to be ruggedized and without loose wires that can be snagged and split, an AugCog system that measures neurophysiological signals must confront the considerable “noise” introduced by motion, sweating, and muscle activity. The preceding chapters covered the means by which these artifacts were addressed for the participants operating in the mobile, multitasking scenarios. In addition, the AugCog program transitioned over the life of the program from using the tethered BioSemi Active Two system with 32 channels of EEG to the wireless ABM six-channel sensor headset. The BioSemi had separate, free wires running from each electrode in an EEG cap to the ribbon cable that connected to the AD (analog-to-digital) box in a backpack. The ABM system had only six channels connected by wires that were integrated and concealed within a mesh cap. The wires led to a short cable bundle that connected to the ABM AD box, which rested flush against the back of the participant’s head and transmitted wirelessly to the mobile processor. The next steps to improve system reliability will involve rigorous testing within dismounted operational environments that will expose the system to increased physical stress and likely introduce new classes of signal artifacts that have not been yet encountered. This would give Honeywell an opportunity to improve the signal processing by isolating and addressing, either with advanced data filtering or physical integration improvements, the new sources of noise.

7.2.2 System Fieldability

Effective integration with Army component systems essentially means efforts need to continue to reduce the hardware, software, computational, and power footprint of the system. Since Phase 2, AugCog has transitioned from a five-desktop, immobile AugCog system to a fully wearable mobile system that relies on only a laptop computer in the shadower’s backpack (see Figure 95). The next step is to incorporate the processing onto the Soldier's on-the-body processor. In addition to the dramatic hardware reduction, the sensing and signal processing requirements are now accomplished by a single, standard laptop. Honeywell will need to continue to streamline to ensure that the sensing system is as small and power-efficient as possible. Furthermore, Honeywell will explore reducing computational requirements by encoding neurophysiological signal processing onto a hardware system that would require less software computation from the wearable computer. Finally, Honeywell will also address potential network protocols that utilize the minimum bandwidth while still transmitting the requisite volume of feedback to provide value to the Army suite of systems. This might also require secure, efficient, and wireless data transmission from the integrated sensors to a conveniently located, miniature hardware signal processor for managing artifacts and spectrally decomposing signal for subsequent classification. Ultimately, a fielded AugCog system will likely consist of advanced sensors integrated with considerable hardware signal processing that are integrated with highly efficient software agents running on the mobile computer for triggering adaptations to the Warfighters’ task environment based on their cognitive state.

Figure 95. Phase 2 (left) and Phase 4 (right) systems.

The next steps to improve fieldability will likely include exploring sensor options that have a reduced footprint. For example, designers will likely consider free-field or minimal-prep EEG electrode-based systems that could be more easily integrated into a helmet liner or embedded within helmet pads.

In addition to investigating more deployable sensors, the Honeywell team will maintain technical coordination with U.S. Army representatives, as well as Army system providers, to align the systems with the most likely configuration of the Army system of systems, such as the wearable computer component. For example, the team will investigate improved cognitive classification by leveraging existing Army Spiral 2 systems such the Vital Sign Detection System (VSDS). The Honeywell AugCog team is currently collaborating on a research initiative to test the reliability and effectiveness of the VSDS in recording physiological data on Soldiers during various physical activities. AugCog will investigate the continued use of the VSDS output (ECG) for cognitive state assessment.

7.2.3 System Form and Function Acceptability

Finally, Honeywell must field an *extremely well-accepted system to ensure use in the battlefield.* User acceptance for an AugCog system includes ease of donning and doffing, comfortable integration with the Advanced Combat Helmet (ACH), and satisfaction of functional expectations. Specifically, the system would need to be seamlessly integrated in the ACH to a degree that Warfighters could simply don their helmet to enable the sensors that are either integrated within the helmet liner or helmet padding—without any adhesives or electrolyte gel. Not only must the sensor-enabled helmet be easy to put on and take off, it should be reasonably comfortable to wear for extended periods. Finally, the AugCog system should deliver value and satisfy functional expectations to justify the addition, however small, of power, weight, and computational requirements. In addition to closing the loop on task adaptations, several Army representatives have expressed an interest in open-loop AugCog applications to allow commanders to evaluate the cognitive combat readiness of their subordinate squads as well as the squad leaders. The next step in addressing these challenges is experimentation with the battle labs environment that will introduce additional form and function requirements. This step will also provide a test environment to do cognitive classification studies with considerably more ecological

validity that should help convince U.S. Army decision makers that the Honeywell AugCog system can detect cognitive states of interest in a relevant environment.

7.3 Lessons Learned

In conducting the ACTE during Phase 4 of the program, several lessons were learned, including:

- The physiological and neurophysiological sensors and sensor system need to be further ruggedized to enable deployment of this capability.
- Thorough advanced signal processing algorithms are essential for use of the measurement of cognitive metrics, particularly to remove or identify noise artifacts in the harsh operational environment.
- There is no one-size-fits-all approach to cognitive state classification. Individualized measurements are necessary for each participant. In addition, due to the nonstationarity of physiological data over time, regular baselines will need to be captured to obtain a high level of classification accuracy.
- The assessment of classification effectiveness will always require evaluation to capture the context of the mission and task, as well as user feedback, as a basis of ground truth information. In addition to a complete understanding of the target environment, thorough interviews with participants and multiple raters of ground truth classification will help minimize any error in cognitive state classification due to poor insight into the cognitive loading requirements of the task environment.

7.4 Conclusions

In conclusion, the Honeywell team believes it was the first ever to demonstrate robust real-time cognitive state classification in the harsh operational military operations in urban terrain (MOUT) environment. Furthermore, the workload classification accuracies obtained in the ACTE at Aberdeen Proving Ground match those of the more pristine laboratory environment, despite the motion, noise, and physical challenges posed by collecting physiological data in the field during real operations. Recent work in sensor deployment and integration to create a mobile ensemble clearly demonstrates the feasibility of this technology for near-ready deployment. The program continues to advance the AugCog capabilities by more fully integrating the cognitive state processing techniques into the information-networked environment. Honeywell looks forward to continuing to meet the needs of Army programs for cognitive state assessment.

This document reports research undertaken at the U.S. Army Natick Soldier Research, Development and Engineering Center, Natick, MA, and has been assigned No. NATICK/TR- 09 / 004 in a series of reports approved for publication.

8 References

Adams, M.J., Tenney, Y.J., and Pew, R.W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors*, *37*(1), 85-104.

Anderson, J.R. (1995). *Cognitive Psychology and Its Implications* (2nd Ed.). New York: Freeman.

APA Monitor (1988). Vincennes: Findings could have helped avert tragedy: Scientists tell Hill panel.

Berka, C., Levendowski, C., Cvetinovic, M.M., Petrovic, M.M., Davis, G., Lumicao, M.N., Zivkovic, V.T., Popovic, M.V., & Olmstead, R. (2004). Real-time analysis of EEG indices of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human Computer Interaction, 17*(2), 151-170.

Blackwell, C. (2003), "Objective Force Warrior: Advanced Technology Demonstration," *Natick Soldier Center*, presentation to Honeywell at the NATICK IPR, October 23.

Breakspear, M., and Terry J. R., (2002). Topographic organization of nonlinear interdependencies in multichannel human EEG. *Neuroimaging,* 16(3 Pt 1):822-35.

Broadbent, D. E. (1958). *Perception and Communication*. New York: Pergamon.

Buller, M.J., Hoyt, R.W., Ames, J., Latzka, W., & Freund, B. (2005). Enhancing Warfighter readiness through physiologic situational awareness—The Warfighter physiological status monitoring—Initial capability. *Proceedings of the 1st International Conference on Augmented Cognition,* Mahwah, NJ: Lawrence Erlbaum Associates.

Burgess, P.W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research*, vol. 63, no. 3-4, pp 279-288.

Cabon, P.., Coblentz, A., Mollard, R., & Fouillot, J.P. (1993). Human vigilance in railway and long-haul flight operations, *Ergonomics,* 36(9): 1019-1033.

Colquhoun, W. P. (1985). Hours of work at sea: Watchkeeping schedules, circadian rhythms and Efficiency. *Ergonomics,* 28(4): 637-653.

Compte, A., Brunel, N., Goldman-Rakic P.S., & Wang X.J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex,* 10(9):910-23.

Cranstoun, S.D., Ombao, H.C., von Sachs, R., Guo, W., & Litt, B. (2002). Time-frequency spectral estimation of multichannel EEG using the Auto-SLEX Method. *IEEE Trans. Biomedical Eng.,* 49(9):988-96.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society,* 39, pp. 1-38.

Dorneich, M.C., Whitlow, S.D., Miller, C.A., and Allen, J.A. (2001). Policy as an interaction method for decision support systems. *Proc. of the 45th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA: HFES.

Dorneich, M.C., Whitlow, S.D., Mathan, S., Ververs, P.M., Pavel, M., & Erdogmus, D., (2005). *DARPA Improving Warfighter Information Intake under Stress - Augmented Cognition: Phase 3 Final Report,* Technical Report for *DARPA Augmented Cognition* Phase 3 under contract DAAD16-03-C-0054, Honeywell Laboratories, December 31, 2005.

Dorneich, M.C., Whitlow, S.D., Ververs, P.M., Mathan, S., Raj, A., Muth, E., Hoover, A., DuRousseau, D., Parra, L., & Sajda, P. (2004b). *DARPA Improving Warfighter Information Intake under Stress - Augmented Cognition: Concept Validation Experiment (CVE) Analysis Report for the Honeywell Team,* Technical Report for *DARPA Augmented Cognition* Phase 2B under contract N66001-01-C-8076, Honeywell Laboratories, December 31, 2004.

Dorneich, M., Whitlow, S., Ververs, P.M., Carciofini, J., & Creaser, J. (2004a). Closing the loop of an adaptive system with cognitive state. *Proceedings of the Human Factors and Ergonomics Society Conference*, Santa Monica, CA: HFES.

Duda, R.O, Hart, R.E., & Stork D.G. *Pattern Classification*, Second Edition. John Wiley & Sons, New York, 2001.

Durlach, P.J. (2004). Army digital systems and vulnerability to change blindness. *Proceedings of the 24th Annual Army Science Conference*, Orlando, FL.

DuRousseau, D.R. (2001). "Method and System for Initiating Activity Based on Sensed Electrophysiological Data." Patent Application Number: PCT/US/50509, Patent Date: Dec. 18, 2001.

DuRousseau, D.R. (2004). Spatial-frequency patterns of cognition. *The AUGCOG Quarterly,* 1(3):10.

DuRousseau, D.R. (2004b), *Multimodal Cognitive Assessment System,* Final Technical Report, DARPA, DAAH01-03-C-R232.

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation, *J. of the American Statistical Association*, 78, 316-331.

Ellis, J. (1996). Prospective memory or the realisation of delayed intentions: A conceptual framework for research. In Brandimonte, M. Einstein, G.O., & McDaniel, M.A. (Eds,), *Prospective Memory: Theory and applications*, 1-22.

Erdogmus, D., Adami, A., Pavel, M., Lan, T., Mathan, S., Whitlow, S., & Dorneich, M. (2005). Cognitive state estimation based on EEG for augmented cognition, *2nd IEEE EMBS International Conference on Neural Engineering*, Arlington VA, March 16-19, 2005.

Freeman, F.G., Mikulka, P.J., Prinzel, L.J., & Scerbo, M.W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking system. *Biological Psychology,* 50, 61-76.

Freeman, W.J., and Skarda, C.A. (1985). Spatial EEG patterns, non-linear dynamics and perception: The neo-Sherringtonian view, *Brain Research Review,* vol. 10, pp. 147-75.

Fuchs, M., Wischmann, H..A., Köhler, Th., &Wagner, M. (1996). The local contribution to the field and the noise induced std. dev. as criteria for the iterative

refinement of current density reconstruction, *Med. & Biol. Eng. & Computing,* 34(2):249-50.

Funk, H., Miller, C., Richardson, J., Johnson, C., & Shackleton, J. (2000). Applying intent-sensitive policy to automated resource allocation: Command, communication and, most importantly, control. *Proceedings of the 5th International Conference on Human Interaction with Complex Systems,* Urbana, IL, pp. 179-183.

Future Force Warrior. (2004). Retrieved September 24, 2004, from http://www.natick.army.mil/ffw/content.htm.

Garavan, H., Ross, T.J., Li, S.-J., & Stein, E.A. (2000). A parametric manipulation of central executive functioning using fMRI. *Cerebral Cortex, 10*, 585-592.

Gevins, A., & Smith, M. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex,* 10(9):829-39.

Gevins, A., & Smith, M. (2003). Neurophysiological measure of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science,* 4(1-2), 113-132.

Gevins, A., & Smith, M.E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex,* 10, 829-839.

Gevins, A., Smith, M.E., McEvoy, L., & Yu, D. (1997). High resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, *7,* 374-385.

Gevins, A.S., Cutillo, B., DuRousseau, D.R., Smith, M.E., et al. (1994). High-resolution evoked potential technology for imaging neural networks of cognition. In Thatcher, R.W., et al. (Eds.), *Functional Neuroimaging: Technical Foundations.* Orlando, FL: Academic Press, Inc., pp. 223-231.

Girolamo, H.J. (2005). Augmented cognition for warfighters: A beta test for future applications. *Proceedings of the 1st International Conference on Augmented Cognition*, Mahwah, NJ: Lawrence Erlbaum Associates.

Gobbele, R., Waberski, T.D., Schmitz, S., Sturm, W., &Buchner, H. (2002) Spatial direction of attention enhances right hemispheric event-related gamma-band synchronization in humans. *Neuroscience Letters,* 327(1):57-60

Gundel, A., & Wilson G.F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography,* 5(1):17-25.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In Hancock, P., & Meshkati, N. (Eds.), *Human Mental Workload*. The Netherlands: Elsevier.

Hockey, G.R.J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. II). New York: Wiley.

Hoover, A., & Muth, E. (2004). A real-time index of vagal activity. *International Journal of Human-Computer Interaction*, *17*(2), 197-210.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kass, S.J., Doyle, M., Raj, A.K., Andrasik, F., & Higgins, J. (2003, April). Intelligent adaptive automation for safer work environments. In J.C. Wallace & G. Chen, *Occupational Health and Safety: Encompassing Personality, Emotion, Teams, and Automation.* Symposium conducted at the Society for Industrial and Organizational Psychology 18th Annual Conference, Orlando, FL.

Kramer, A. (1991). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), Multiple Task Performance (pp. 279-328). London: Taylor and Francis.

Kremper, A., Schanze, T., & Eckhorn R. (2002) Classification of neural signals by a generalized correlation classifier based on radial basis functions. *Journal of Neuroscience Methods,* 116(2002):179-87.

Levy ,R., & Goldman-Rakic P.S. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. *Exp. Brain Res.*, 133(1):23-32.

Makeig, S., & Jung, T-P. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *NeuroReport, 7*(1), 213-216.

Makeig, S., Enghoff, S., Jung, T., & Sejnowski, T. (2000). A natural basis for efficient brain-actuated control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 8(2):208-11.

Mathan, S., Mazaeva, N., Whitlow, S., Adami, A., Erdogmus, D., Lan, T., & Pavel, M. (2005). Sensor-based cognitive state assessment in a mobile environment. *Proceedings of the 1st International Conference on Augmented Cognition*, Mahwah, NJ: Lawrence Erlbaum Associates.

Mikulka, P.J., Scerbo, M.W., & Freeman, F.G. (2002). Effects of a biocybernetic system on vigilance performance. *Human Factors*, 44(4), 654-664.

Montain, S.J., Sawka, M.N., & Wenger, C.B. (2001). Hyponatremia associated with exercise: Risk and pathogenesis. *Exercise Sports Science Review*, *29*, 113-117.

Norman, D.A. (1990). The problem with “automation”: Inappropriate feedback and interaction, not “over-automation.” *Philosophical Transactions of the Royal Society of London*, B327, 585-593.

O’Hanlon, J.F., & Beatty, J. (1977). Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance. In R.R. Mackie (Ed.), *Vigilance: Theory, Operational Performance and Physiological Correlates* (pp. 189-201). New York. Plenum.

Parasuraman, R., & Davies, D.R. (1984). *Varieties of Attention*. New York: Academic Press.

Parmentola, J.A. (2004). Army transformation: Paradigm-shifting capabilities through biotechnology, *The Bridge*, 34(3), The National Academy of Engineering of the

National Academies. Retrieved January 28, 2005, from http://www.nae.edu/NAE/bridgecom.nsf.

Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., & Sajda, P. (2002). Linear spatial integration for single-trial detection in encephalography, *NeuroImage,* vol. 7, no. 1, 2002.

Parra, L., Spence, C., Gerson, A., & Sajda, P. (2003). Response error correction - A demonstration of improved human-machine performance using real-time EEG monitoring, *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 11, no. 2, pp. 173 -177, June 2003.

Parzen, E. (1967). On estimation of a probability density function and mode. *Time Series Analysis Papers*, San Diego: Holden-Day.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220-244.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology* 37(2):127-52.

Pilcher, J., & Muth, E. (2003). *Anlaysis of Data from Sleep Deprivation Study,* Technical Report.

Pope, A.T., Bogart, E.H., & Bartolome, D.S. (1995). Biocybernetic system validates index of operator engagement in automated task. *Biological Psychology, 40*, 187-195.

Popivanov, D, & Mineva, A. (1999). Testing procedures for non-stationarity and non-linearity in physiological signals. *Mathematical Biosciences*, 157(1-2), 303-20.

Prinzel III, L.J., Scerbo, M.W., Freeman, F.G., & Mikulka, P.J. (1997). Behavioral and physiological correlates of a bio-cybernetic, closed-loop system for adaptive automation. In M. Mouloua & J.M. Koonce (Eds.), *Human-Automation Interaction: Research and Practice* (pp. 66-75). Mahwah, NJ: Lawrence Erlbaum Associates.

Prinzel III, L.J., Hadley, G., Freeman, F.G., & Mikulka, P.J. (1999). Behavioral, subjective, and psychophysiological correlates of various schedules of short-cycle automation. In M. Scerbo & K. Krahl (Eds.), *Automation Technology and Human Performance: Current Research and Trends*. Mahwah, NJ: Lawrence Erlbaum Associates.

Raj, A.K., Kass, S.J., & Perry, J.F. (2000). Vibrotactile displays for improving spatial awareness. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting.* Santa Monica, CA: HFES.

Raj, A.K., Perry, J.F., Abraham, L.J., & Rupert A.H. (2003). Tactile interfaces for decision making support under high workload conditions. *Aerospace Medical Association 74th Annual Scientific Meeting*, San Antonio, TX.

Raley, C., Stripling, R., Schmorrow, D., Patrey, J., & Kruse, A. (2004). Augmented cognition overview: Improving information intake under stress. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: HFES.

Rowe, J., Friston, K., Frackowiak, R., & Passingham R. (2002). Attention to action: Specific modulation of corticocortical interactions in humans. *Neuroimaging,* 17(2):988.

Russell, C.A., & Gustafson, S.G. (2001). *Selecting Salient Features of Psychophysiological Measures,* Air Force Research Laboratory Technical Report (AFRL-HE-WP-TR-2001-0136).

Sajda, P., Gerson, A., & Parra, L. (2003). Spatial signatures of visual object recognition events learned from single-trial analysis of EEG. *IEEE Engineering in Medicine and Biology Annual Meeting*, Cancun, Mexico.

Sarter, N.B., Woods, D.D., & Billings, C.E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd edition) (pp. 1926-1943). New York: Wiley.

Schanze, T., & Eckhorn R. (1997). Phase correlation among rhythms present at different frequencies: Spectral methods, application to microelectrode recordings from visual cortex and functional implications. *International Journal of Psychophysiology,* 26(1997): 171-89.

Schmorrow, D.D., & Kruse, A.A. (2002). Improving human performance through advanced cognitive system technology. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC'02)*, Orlando, FL.

Schneiderman, H., & Kanade, T. (2004). Object recognition using statistical modeling. Available online at http://www.ri.cmu.edu/projects/project_320.html.

Steinman, A.M. (1987). Adverse effects of heat and cold on military operations: History and current solutions. *Military Medicine,* 152, 389-392.

Stone, P. (2003). Cebrowski sketches the face of transformation. Retrieved January 28, 2005, from http://www.defenselink.mil/news/Dec2003/n12292003_200312291.html.

Takahashi, N. (2006). Efficient learning algorithms for support vector machines. Available online at http://www-kairo.csce.kyushu-u.ac.jp/~norikazu/ research.en.html.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature,* 381, 520-2.

Treisman, A.M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology,* 77, 206-219.

U.S. Army Aberdeen Test Center. (2007). Online at: http://www.atc.army.mil/pages/aboutATC/ aboutus.html

U.S. Navy. (1988). *Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988*. Washington, DC: Department of Defense Investigation Report.

U.S. Army. (2003), *The Objective Force Warrior: The Art of the Possible . . . a Vision*. Available online at http://www.ornl.gov/sci/nsd/pdf/OFW_composite_vision.pdf.

Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer-Verlag.

Ververs, P.M., Whitlow, S.W., Dorneich, M.C., & Rye, J. (2003). *Sensor Suite Development for Cognitive Optimization,* Final Technical Report, Honeywell

Laboratories, submitted to DARPA under contract N66001-01-C-8076, Task 6.1 Augmented Cognition, Mixed-Initiative Control of Automata Program.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70-73.

Whitlow, S.D., Dorneich, M.C., Ververs, P.M., Raj, A., DuRousseau, D., Parra, L., Sajda, P., & Muth, E. (2004). *Phase 2A Concept Validation Experiment Final Report,* Technical Report for *DARPA Augmented Cognition* Phase 2A under contract N66001-01-C-8076, Honeywell Laboratories, January 30, 2004.

Wickens, C.D., Heffley, E., Kramer, A., & Donchin, E. (1980). The event-related brain potential as an index of attention allocation in complex displays. *Proceedings of the 24th Annual Meeting of the Human Factors Society*. Santa Monica, CA: HFES.

Widrow, B., & Hoff, M.E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, pp. 96-104.

Wolfowitz, P. (2002, April 9). The Imperative for Transformation. Prepared Statement for the Senate Armed Services Committee Hearing on Military Transformation. United States Department of Defense, Washington, DC. Retrieved February 28, 2007, from http://www.defenselink.mil/speeches/2002/s20020409-depsecdef2.html.

Wood, R., Maraj, B., Lee, C.M., & Reyes, R. (2001). Short-term heart rate variability during a cognitive challenge in young and older adults. *Age and Ageing,* 2002; 31: 131-135.

INTENTIALLY LEFT BLANK

Appendix A
List of Acronyms

Acronym	Description
ABM	Advanced Brain Monitoring, Inc.
ACTE	Augmented Cognition Test Event
AD	Analog-to-digital (converter)
AM	Augmentation Manager
ANOVA	Analysis of Variance
ATC	Aberdeen Test Center
AugCog	Augmented Cognition
BLUFOR	Friendly (Blue) Force
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance
CCNY	City College of New York
CGF	Computer Generated Force
CLIP	Closed-Loop Integrated Prototype
CMU	Carnegie Mellon University
CO	Company Commander
CSA	Cognitive State Assessor
CSP	Cognitive State Profile
CVE	Concept Validation Experiment
CWA	Cognitive Workload Assessor
DARPA	Defense Advanced Research Projects Agency
DRM	Dead Reckoning Module
DoD	Department of Defense
ECG	Electrocardiogram
EDR	Electrodermal Response
EEG	Electroencephalogram
EMG	Electromyogram

EOG	Electro-oculogram
FFT	Fast-Fourier Transform
FFW	Future Force Warrior
fNIR	functional Near Infrared
GPS	Global Positioning System
GSR	Galvanic Skin Response
GUI	Graphical User Interface
HUD	Head Up Display
HMI	Human-Machine Interface
IBI	Interbeat Interval
IED	Improvised Explosive Device
IFF	Identify Friend or Foe
IHMC	Institute for Human and Machine Cognition
JDFE	Joint Distributed Freeplay Event
MOUT	Mobile Operations in Urban Terrain
MRTFB	Major Range and Test Facility Base
NASA	National Aeronautics and Space Administration
NCNG	North Carolina National Guard
NSRDEC	Natick Soldier Research, Development and Engineering Center
O/C	Observer/Controller
OPFOR	Opposing Force
PDA	Personal Digital Assistant
PL	Platoon Leader
PSD	Power Spectral Density
PSG	Platoon Sergeant
RMS	Root Mean Squared
ROC	Receiver Operating Characteristic
SA	Situation Awareness
SL1, SL2, …	Squad Leader 1, Squad Leader 2 (and so on)
TLX	Task Load Index
TSAS	Tactile Situation Awareness System

UAV	Unmanned Air Vehicle
UTA	Utility Task Analysis
VE	Virtual Environment
VOG	Video Pupilometry
VSDS	Vital Signs Detection System
XLI	eXecutive Load Index (gauge)

INTENTIONALLY LEFT BLANK

Appendix B
Phase 2a CVE Qualitative Feedback

Each participant in the Phase 2a Concept Validation Experiment (CVE) was given a post-experiment questionnaire consisting of two parts: rating scales and short-answer questions.

B.1 Ratings

Table B-1 gives the rating scale averages and comments.

Table B- 1. Rating scale averages and comments.

Question	Average (Std. Dev.)	Comments
1) I found it easy to remember the assigned route for each trial	6.08 (0.90)	Seeing same location in a different route was confusing (S8). I think only two practices are necessary (S12).
2) I found it easy to identify my team members from the enemy Soldiers	5.67 (1.15)	Enough distinction provided between the two (S8). The color was hard to identify from far off (S12).
3) I felt it was more difficult to engage multiple enemy Soldiers at once, rather than one at a time.	4.83 (1.53)	Focus is drawn to one while others are free to return fire (S8). It is not more difficult, just takes different strategies (S14).
4) I found it difficult to listen and respond to messages that required an answer.	5.25 (1.14)	Especially immediately before a firefight starts (S8). Problems while engaging the enemy during a message (S9). I am not used to doing that while playing (S10). There was too much going on to do that (S11). True when engaging the enemies (S12). Sometimes I could only hear part of the question (S13)
5) I found it difficult to listen to and remember messages that did not require an answer	6.17 (1.19)	Because they were “not as important” (S8). Same as before (There was too much going on to do that) (S11). It was hard to remember specific details, but not the message (S14). Some were hard to understand (S15).
6) I found controls for navigating and shooting easy to learn.	6.81 (0.39)	Common configuration (S8). I knew the controls before the experiment (S12)
7) The text in the message window was easy to read.	5.09 (1.76)	No messages presented (S8). The messages did not stand out well to catch the eye (S9). Easy to read, but hard to see who it came from (S12)

B.2 Short-Answer Questions

This section lists the 11 short answer questions and associated responses.

1) Explain what made it easier or harder for you to understand, remember, or respond to messages.

- Spaced repetition (S5).
- It is easier to understand a message if there is a tone before it to prepare you for the following message (S6).
- The different tones made responding and remembering easier (S7).
- Firefight situations while receiving message and messages repeating before a chance to answer or when already answered made it harder (S8).
- It was hard to remember non-important messages because of the concentration on going the correct direction, engaging enemies, and the text box (S9).
- It was easy to respond to yes/no questions rather than more complex questions (S10).
- It was harder to remember when there was a lot going on, and the voice of the commander sounded like he was gargling screws (S11).
- The messages that were hard to remember were those beginning with names of people or squads (S12).
- It made it easier for me to understand and respond to messages when they were not being prioritized by beeps. When the priority sound was being played, I found that I did not pay attention to messages without a priority. Also, I tended to miss messages while I was engaged in combat (S13).
- The more familiar I got with the terms, the easier it got to remember (S14).
- Some of the voices were hard to understand (S15).
- Enemy fire made everything much harder (S16).

2) Did you find it easier to complete trials toward the end of the study? Why, or Why not?

Yes = 100%, No = 0%

- Routes were more familiar. Knew to pay closer attention to messages.
- I began to get used to the fact that enemies do not move. Also remembered the route.
- I found that I was rushing so I started slowing down and listening more.
- May have been due to familiarity with the situation.
- I was more familiar with the route to take the objective.
- Was more familiar.
- I learned the routes and controls.
- The pattern and location of enemies seemed to be repeated. I also began to listen to certain aspects of the messages that I thought I would be asked to recall.
- I was more familiar with the layout of the map. Also, in the end messages were played without beeps, making me more responsive to them.
- Familiar with the terms used.

- It was easier to concentrate without the text box.
- Knew what to expect.

3) Completing the final trial in a standing position was significantly more difficult than the other trials

True = 58%, False = 42%

- So much effort was required to maintain leg movement that the other things were difficult to keep up. It may be easier if I were actually walking, or if game environment responded directly to my movement.
- Harder to keep your mouse straight.
- While it added to the overall difficulty, it did not require a lot of thought to continue running.
- Impaired input to keyboard and mouse.
- I was also trying to maintain a steady jog while stopping, turning corners, and engaging the enemy.
- Am not used to having to move while playing.
- It was hard for me to keep my hand on the controls while walking.
- I was concentrating on keeping my balance rather than the enemy or the messages.
- The only trouble with the standing trial was concentrating on marching. This did not significantly detract form the trial.
- It was only slightly difficult to keep in step.
- It was a little more difficult, but not a whole lot.
- Not sure. Just seemed about the same.

4) Did you adopt a strategy for handling messages while performing the trials? Please Explain.

Yes = 75%, No = 25%

- Repeat info to myself.
- Took keywords from message.
- I tried repeating several times in my head the message I received.
- When I received important messages, I tried to repeat the message to myself several times.
- When a message would begin to play I would slow down a bit.
- Explained in Question 2.
- For the beeps, I tended to ignore messages without a priority. For the other trials, I tended to consider messages more and place more effort into my response.
- Mentally repeating them.
- I tried to associate each voice with a face. Then I tried to make a generalization about what that face usually asked me.
- Handling messages was not that hard. Remembering them was the hard part.

5) Did you emphasize one task over another?

Yes = 82%, No = 18%

- Nav was no trouble. I tended to concentrate more on safety (i.e., killing hostiles).
- Navigating. It seemed to be the most important part of the mission.
- Emphasized completion of task in a timely manner and eliminating enemies.
- I concentrated mainly on not getting shot, then direction to objective.
- Navigating. I wanted to achieve that over all else.
- When navigating and Identify Friend of Foe (IFF) became easier, I began to listen to messages more intently.
- I tended to focus more on navigation and identifying than on the messages during the beep trials. During the trials without the beeps, I focused more on the messages. This is also because I was more familiar with my route during these trials.
- Not really.
- I tried to listen to the messages. I thought everything else was easy.
- Listening to messages. That was the hardest part.

6) Did you read text messages during trial? If so, when during the trial?

Yes = 67%, No = 33%

- W/ frequent glances, and at the end before declaring.
- Very little, after I clear an area.
- Rarely, I was usually occupied with navigating and identifying enemies.
- No text messages provided.
- I tried to read them as they appeared. Sometimes I would try to read them as I reached the objective.
- When I was sure I had time.
- In the beginning.
- During transit when I wasn't expecting enemy contact.
- When there was no threat.
- When no enemies were around.
- At the end.

7 and 8) Rank (on a five-point scale) the difficulty of each task, and your perceived performance on each task.(See Table B- 1 for average of responses.

Table B- 2. Participant's difficulty and performance ratings.

Task	Difficulty: Average (StD)	Performance: Average (StD)
Navigating to Objective	2.08 (1.24)	4.58 (0.79)
Identifying Friend or Foe	2.25 (1.06)	4.22 (0.65)
Managing Communications	4.33 (0.78)	2.17 (0.39)

9) Did you feel that the tones presented before certain messages helped you attend to and understand those messages better than messages without tones?

Yes = 67%, No = 33%

- I found that verbal cues during or the end of a message helped more than tones. Tones may take on more meaning with training.
- It prepared me for the message.
- It helped me decide how much attention to pay if I was busy.
- Notification prepared me to receive a message.
- The tones provided a unique sound away from gun shots and footsteps to alert me of a message.
- The tones became very annoying.
- I allotted more concentration to the messages following tones.
- Because those messages had a priority, they tended to make me listen to only those messages. I was still confused as much with those messages as ones without the beep.
- The messages with tones made me listen more and try to remember.
- I heard the tones but the pitch of the tone did not make me concentrate more.
- It prepared me for the message.

10) Could you tell the difference between the four tones used in the study?

Yes = 75%, No = 25%

- They were very clear,
- Not all of them. The high priority tones were differentiable, but others I missed.
- I had no problem distinguishing the tones.
- Not really. I was concentrating more on completing the objective and trying to listen to the message after the tone.
- All the tones were very different.
- I could decipher between only 2. Urgent and Medium.
- I was able to recognize and use the tones, but I only remember 3.
- They were different enough.
- I was trying to concentrate on everything else.

11) Other comments

- The eye tracker needs more padding.
- The uncomfortable equipment could affect results.
- Temperature might have been a factor. No complaints.
- It is hard to hear over gunfire and some messages from command were hard to understand.
- Was interesting.

INTENTIONALLY LEFT BLANK

Appendix C
Phase 2b CLIP Configuration

C.1 Hardware Configuration for CLIP at IHMC CVE

The Phase 2b CVE CLIP consisted of a test participant station that included a keyboard, a mouse to control the Virtual Environment (VE) projected from an overhead projector onto a flat screen approximately 7 feet in front of the participant, a Tablet PC with stylus and mouse, and an instrumented helmet carrying a microphone, ear bud stereo headphones, and the IScan eye-tracking cameras and dichroic mirrors.

The ActiveTwo EEG system was placed at head level immediately behind the participant. The participant wore, under the helmet, a 34-scalp-electrode EEG cap. The participants had three electrodes taped near their eyes, one electrode on the mastoid for HEOG/VEOG and reference and two electrodes taped to their chest at V1 and V6. In addition, a blood volume pulse plethysmograph sensor was strapped to the right-hand fourth phalange, two galvanic skin conductance sensors were taped to the left-hand second and third phalanges, a temperature sensor was taped the dorsum of the right hand, and a respiration monitor strap was placed about the thorax. The 10 Cardiax ECG electrodes were placed in standard configuration on the chest and limbs for 12-lead ECG analysis. Each test participant also wore the Tactile Situation Awareness System (TSAS) belt to provide navigation information.

The ActiveTwo and Cardiax devices connected to PC workstations via USB, and the ActiveTwo also was connected to the sound application or Visual Calibration agent via a parallel port A-B switch (selected by the test operator). The IScan cameras and illuminators connected to the IScan PCI card via a video synch driver/power supply. Audio was mixed and recorded on the audio channel of the Hi8 VCR, which also recorded a video image of the VE projection screen.

Physiologic data were captured by the PC workstations and transferred between agents as needed. The test operator was positioned to monitor the agent and system displays of the PC workstations and could launch and adjust all agents as necessary. Data logging was performed by all agents (sensors and CWA agents and applications) locally in binary form, and the resulting files were collected and posted *post hoc*.

C.1.1 Workstation Configuration

Figure C- 1 depicts the workstation configuration of the IHMC CVE.

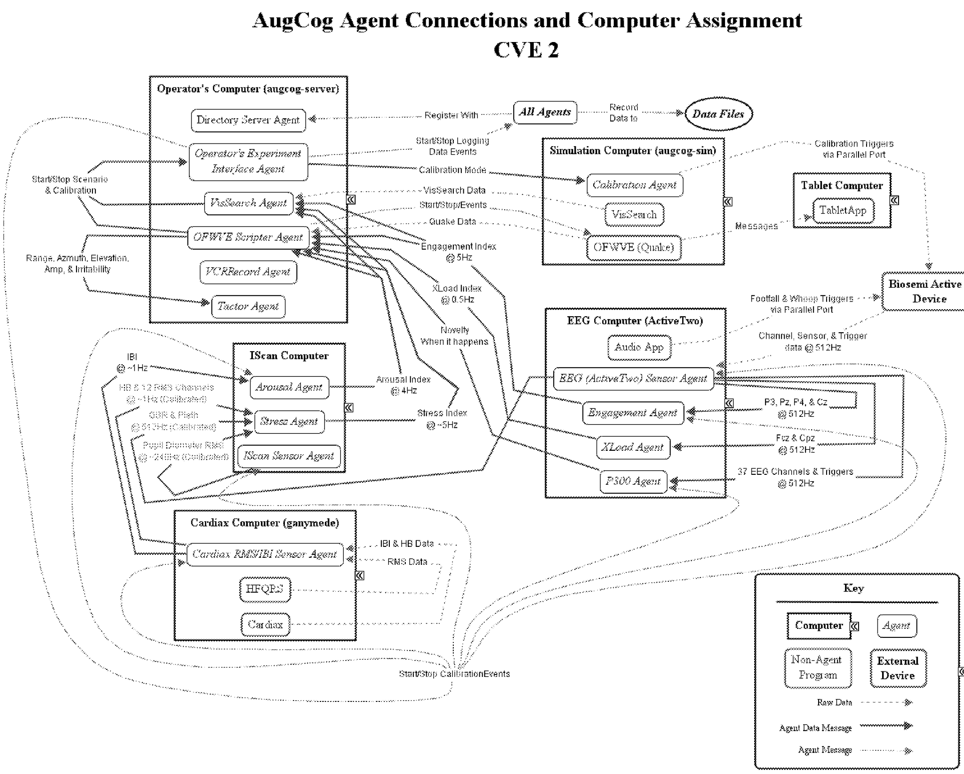


Figure C- 1. Agent-based architecture (IHMC Phase 2b CVE AugCog implementation).

Each box represents an agent that took input from other agents, non-agent software, or physical hardware. Seven PC workstations were employed for this CVE:

1. *Server* (2.2 GHz P4, Windows XP) running Agent Launcher, Time Server, Directory Service, Experimental Console, Console Helper, Augmentation Manager, Tactor Agent, and the VCR Record Agent
2. Simulator (AMD Athlon 2200 nForce2, Windows XP) running Agent Launcher, Time Fixer Client, FFW VE, and the Visual Calibration Agent
3. *Cardiax* (2.0 GHz P4, Windows 2000) running Agent Launcher, Time Fixer Client, CardioSoft, Cardiax Agent, and the CrxHFQRS Agent
4. *EEG PC* (AMD Athlon 2800, Soundblaster, Windows XP) running Agent Launcher, Time Fixer Client, ActiveTwo Agent, Engagement Agent, P300 Agent, HBXload Agent, and the Sound App
5. *IScan PC* (AMD Athlon 3000, nForce2, Windows XP) running Agent Launcher, Time Fixer Client, IScan Agent, Arousal Meter Agent, and the Stress Agent
6. *Tablet PC* (Fujitsu Tablet PC, Windows XP Tablet Edition) running the Tablet App
7. *Tactor Driver* (400-MHz P2, QNX) running the TSAS Driver

All workstations were connected via an isolated, wired local area network using two 100/10 BaseT network switches. All workstations used 100 BaseT except the Tactor Driver, which is limited to 10 BaseT.

C.1.2 Sensor/Gauge System Setup

Each gauge was connected to specific hardware. Namely, the ActiveTwo (BioSemi, Netherlands) EEG device connected to the HBXload gauge (CPz and FPz), engagement gauge (P3, P4, Cz, Pz), and P300 novelty detector (all scalp electrodes and EOGs, plus trigger values from the soundPlayer app). In addition, ActiveTwo connected to the Stress Gauge, passing Galvanic Skin Response (GSR) and Pleth, which were recorded but not included in the calculation of stress. Pupil diameter was supplied to the Stress Gauge via an IScan binocular near-infrared high-speed (240-Hz) eye tracker (IScan, Inc., Burlington, MA). The Stress Gauge received heart rate directly from the Cardiosoft driver for the high-speed (1000-Hz) ECG device (Cardiax, Budapest, Hungary) and HFQRS Root Mean Squared (RMS) data from the HFQRS algorithm developed by Dr. Todd Schlegel at NASA JSC. The Arousal Meter also received uncorrected IBI from the Cardiax Cardiosoft system.

C.1.3 Cognitive State Gauges

C.1.3.1 Engagement Gauge

The Engagement Index was an indicator of alertness. It used a ratio of EEG power bands, beta/(alpha + theta). Research has shown a direct relationship between beta and alertness and an indirect relationship between alpha and theta and alertness (Mikulka, et al., 2002; O'Hanlon & Beatty, 1977). Freeman, Mikulka, Prinzel, and Scerbo (1999) have shown the Engagement Index to be a valid measure of an operator's engagement in the task set. Anything above zero indicated higher than normal engagement and anything below zero indicated lower then normal engagement.

The Engagement Index was measured using the power in three separate frequency bands: theta (4-8 Hz), alpha (8-13 Hz), and beta 13-22 Hz. Frequency analysis was performed using the Fast-Fourier Transform (FFT). The band powers were computed from the raw EEG data every X seconds in a window of Y seconds. The average power in each band is further averaged over the relevant electrodes (X,X,Z). Consistent with Freeman, et al.'s (1999) work, EEG data were recorded from sites Cz, Pz, P3, and P4 with a ground site midway between Fpz and Fz. The Engagement Index (beta/ (alpha + theta)) was calculated from a running average of powers for different EEG frequency bands (Prinzel, et al., 1999).

The engagement gauge was initialized with the number of total input channels (required), the number of samples (required), and the sampling frequency (required). The engagement gauge took as input data four EEG channels (P3, P4, Cz, Fz) from the ActiveTwoAgent (512Hz EEG) at 5 Hz (2560 reads/packet). The Engagement gauge outputted six channels at 2 Hz: (1) Engagement Index, (2) y1, (3) y2, (3) y3, (4) y4, and (5) Index_cal.

The engagement algorithm was coded in MATLAB. The Java Agent used a C wrapper to access the MATLAB algorithm. The engagement gauge did its processing on the EEG data based on some predefined bands to generate an index value. These bands were tailored by data created by the HBXload agent. This agent generated bands specific to the individual and saved them to two files. The engagement gauge looked for these files on startup, and if available, overrode the default bands with the ones determined by HBXload.

Frequency Subband Powers: CCNY provided an online block-processing algorithm to compute subband powers implemented in MATLAB. It was configured to compute powers in arbitrary frequency bands and over arbitrary time windows. IHMC configured it to compute the Engagement Index and to estimate 60-Hz power as an indicator for contamination of the EEG signal with inductive environmental noise.

C.1.3.2 Arousal Gauge

The Arousal Meter (AM) was a real-time cardiac-based measure derived from IBIs to status the activity of the autonomic nervous system. The heart responded to changes in the parasympathetic nervous system (PNS), which was responsible for returning an individual person to a resting state, and the sympathetic nervous system (SNS), which was responsible for the “fight or flight” arousal state. High arousal was typically characterized by increases in heart rate, PNS withdrawal, and SNS activation. The AM used the PNS subcomponent of ANS to status an individual’s level of arousal (Hoover & Muth, 2004).

Anything above zero indicated higher than normal arousal, and anything below zero indicated lower than normal arousal. A three-lead ECG was used to detect R-spikes and derive millisecond resolution IBIs that were then re-sampled at 4 Hz. An FFT was computed for 16 seconds, 32 seconds, or 64 seconds worth of IBIs. A sliding window was established such that a new FFT was computed every 0.25 second. When the FFT was computed, the high-frequency peak (maximum power between 9 and 30 cycles per minute) was identified, and the power at that peak, termed respiratory sinus arrhythmia (RSA), was stored. Once one minute’s worth of FFT results are stored, the AM began to generate a standardized arousal which is computed every 0.25 second using a z- log-normal score standardization and the running mean and standard deviation of the RSA values.

For the IHMC CVE, a proprietary version of the AM (version 2.3) that employed a Java wrapper, as well as customized smoothing of the output, was used. The smoothing process used a Kalman filter. It works by assuming that the observed values are a noisy version of the real values, and attempted to predict the true values based on real measurements. A coefficient input into the Kalman filter algorithm determined the degree of smoothing used. Practically, using this filter caused a slight lag in the output of the arousal algorithm. This latency varied with the degree of smoothing employed. In the version of smoothing coefficient used for the IHMC CVE, this latency was approximately 15 seconds. It was also important to note that this proprietary version of the AM utilized IBIs generated from hardware other than the recommended EZ-IBI by UFI (Morro Bay, CA). For the CMU CVE, the standard desktop AM (version 2.2) was used

with the recommended EZ-IBI hardware for IBI generation. Both versions of the AM employed a rudimentary IBI error detection and correction algorithm which corrected isolated IBIs that were either split due to a false trigger or combined due to a missed trigger. In testing, this algorithm was quite effective in stationary participants with the EZ-IBI. Its usefulness with other hardware is unknown. Further, its ability to detect and correct IBI errors in moving participants was quite limited. It was known that even small numbers of IBI errors (one per minute) greatly affected the calculations that drive the AM. Hence, error-free data was the ultimate goal. A sophisticated error detection/ correction algorithm was under development with the goal of increasing the usability of IBI data in a moving participant and accepting IBI input from any standard IBI generating hardware. Nonetheless, this algorithm was not available for the current CVEs, and due to hardware differences and the lack of familiarity of the Clemson team with the hardware used at IHMC, it was difficult to validate the quality of the IBIs from the IHMC data. The IBIs were verified, but certain atypical IBI series could not be clearly identified as artifact or usable data. Hence, in the CVEs, it was likely that IBI artifact during some trials impacted the real-time arousal calculation and decreased the utility of mitigation strategies that were based on the AM. This is discussed further in Chapter 4.

The Arousal Meter was initialized with the number of samples to generate a power set, the number of power sets required for average to find peak in the 10- to 30-CPM range, and the IBI Sample Rate. The AM took as input data one IBI channel from the IBI Agent at a rate of 1 Hz. The AM output the arousal index at 4 Hz.

The arousal algorithm was coded in C. It simply took in IBI values and output arousal index values. The gauge needed to run for 15 minutes to calibrate. It created a file that tracked the current calibration numbers as well as how much time the gauge had been calibrating. If the gauge was shut down and restarted, it would look for this file and reloaded the three calibration variables, eliminating the need to recalibrate the gauge.

Based on pre-CVE data and Clemson's experience with the AM, thresholds were set to recommend mitigation. The AM outputted normalized data that were comparable between participants. The standardized data were interpreted as a z-score from a normal distribution with a mean of zero and a standard deviation of one. Mitigation was accomplished as follows: If arousal was between 0+ and – 0.5, it was recommended that the current mitigation strategy not be changed; if arousal moved above 0.5, workload/arousal was considered high and appropriate mitigation was recommended; and if arousal fell below –0.5, workload/arousal was considered low and appropriate mitigation was recommended.

To achieve a true measure of when the AM indicated a significant change in arousal, two things must be accomplished. First, a baseline must be established such that the data standardization procedures have enough data so that the mean and standard deviation statistics have stabilized. In testing at Clemson, it has been shown that these statistics stabilize after 15-20 minutes of data. Hence, for the CVE a minimum of 15 minutes of baseline data were required. Second, change in arousal must be tracked such that any noise is eliminated, and only significant changes are indicated. There currently was no solution to this problem. To begin to address this, an untested smoothing feature was

introduced at the IHMC CVE. This smoothing feature was not implemented at the CMU CVE. This smoothing was only one approach, and Clemson is currently working to establish guidelines for exactly what constitutes a significant change in arousal such that mitigation should be turned on or off. Hence, the driving of mitigation in the current study must be considered a preliminary test of how this will work, with future testing needed once better guidelines are established.

C.1.3.3 Stress Gauge

The composite Stress Gauge used three main inputs: heart rate (HR), pupil diameter (PD), and microvolt cardiac QRS waveform root mean square (RMS) amplitude (HFQRS) to determine stress during individual trials. Heart rate was determined using an R-R interval detector in hardware by the Cardiax PC ECG device running CardioSoft software, which returned it to the Stress Gauge Agent. HFQRS used the Cardiax R-R interval to determine a standard template for an individual's QRS (millivolt) ECG waveform. This was based on a moving window at 70 highly correlated beats. The difference between this template and each successive beat that cross-correlated sufficiently in 10 or more leads resulted in the residual microvolt electrical activity within the QRS. The RMS of the beat was calculated for each lead and the 12 HFQRS RMS values were z-scored, then averaged and returned to the Stress Gauge agent.

PD used pupilometry from 240-Hz near-infrared videography of both eyes and employed an intelligent pupil tracking algorithm to remove shadows and eyelash artifact, and returned PD for both eyes. The left and right eye PDs were averaged and returned to the Stress Gauge agent. The HFQRS RMS was weighted by –0.2 (decreased RMS is inversely related to stress) of the final score, PD by 0.5, and HR by 0.3, if all three channels were online. If any channel was missing, the weights were adjusted proportionally and the certainty was adjusted to indicate this. Galvanic Skin Response (GSR) and Blood Volume Pulse BVP (Pleth) were recorded, but not included due to excessive noise and drift. EMG from the trapezius muscle was deleted from the gauge in this CVE due to excessive activity (likely due to the weight of the instrumented helmet) and poor correlation in the Phase 1 and 2a CVEs.

The Stress Gauge indicated changes in physiologic stress associated with cognitive tasking. Values above zero indicated increasing stress and values below zero indicated decreasing stress.

The Stress Gauge was initialized with the total input channels, weighting for each channel, and the sampling frequency. The Stress Gauge agent took as input heart rate via Cardiax (millisecond-resolution R-R interval) resampled to 5 Hz, CrxHFQRS resampled at 5 Hz (HFQRS RMS in I, II, III, aVL, aVR, aVF, V1, V2, V3, V4, V5, V6), and pupil diameter of both eyes from IScan (240 Hz) at 5 Hz (1200 reads/packet). The Stress Gauge agent outputted the stress index at 5 Hz.

The Stress Gauge was coded in Java and used the following formula where PD was z-scored pupil diameter and HR is scaled (–1 = 40 to +1 = 120):

$$\frac{0.5\text{x(Right PD RMS+Left PD RMS)}}{2} + 0.3\text{xHR-}0.2\text{x(HFQRS RMS)}$$

where HFQRS RMS was equal to the average of the z-scored HFQRS RMS channels:

$$\frac{\text{I+II+III+aVR+aVL+aVF+V1+V2+V3+V4+V5+V6}}{12}$$

The PD was z-scored during the pupil stare portion of the visual calibration routine, and the HFQRS was normalized during the P300 calibration routine. Because each input to the gauge was normalized to –1 to +1, the inputs could be weighted and added to result in a final gauge output of –1 to +1. If any of the components disconnected, the remaining were proportionally reweighted, and the certainty was adjusted by the recalculated weights as well.

C.1.3.4 P300 Novelty Detector Gauge

The P300 gauge indicated if there was a reaction to a task-relevant novel event. Anything above zero indicated that there was a reaction, and anything below zero indicated there was no reaction. The P300 gauge measured the strength of the EEG-evoked responses following an alert tone. The detector was adapted for every user during a calibration phase. During calibration, the user heard approximately 30-40 alert tones during a low-workload task. The detector was optimized to differentiate the EEG response evoked by the alert tones from that activity evoked by a frequent auditory stimulus of no significance (Soldier footfall sounds were used).

The algorithms for P300 detection had three components: eye calibration, P300 calibration/detection, and data preprocessing. These algorithms process the 37 EEG channels recorded with a BioSemi ActiveTwo device. The data was processed at 512 Hz

Eye calibration used the EEG activity recorded during a 30- to 40-second eye movement sequence during which the participant followed a cross on the screen and blinked repeatedly during a predetermined time. This data was processed by an algorithm that generated three vectors indicating the 3-D subspace of the 37-dimensional EEG data that best described eye blinks and horizontal and vertical eye motion.

The preprocessing algorithms included 60-Hz and 120-Hz filters, a DC drift filter, as well as eye-blink and eye-motion subtraction. It was implemented as an online block-processing algorithm. The algorithm was initialized with the information determined during eye calibration, whenever available.

The P300 detection gauge was trained during a calibration run, and the parameters of the gauge were saved to file. These detector parameters were then used to report a P300 output for each audio alert event. In essence, after preprocessing, the 37-dimensional EEG data was projected onto the orientation that best discriminates between EEG responses evoked by footfall sounds and responses evoked by audio alerts. The detector measured the average evoked response during 300-400 milliseconds following the alert.

The timing of the alert indicated by an event marker channel of the ActiveTwo (BioSemi) device. That signal, in turn, was communicated via a parallel port connection by the application that produced the alert tones. This mechanism bypassed the messaging system of the IHMC agent architecture to ensure millisecond accuracy.

The P300 agent was initialized with the number of total input channels (required), sampling frequency (required), downsampling frequency (default is fsref = fs), positive samples (default pos = 1), negative samples (default neg = 1), cost of false negative (default c1 = 1), and the cost of false positive (default c2 = 1). The P300 agent took as input 37 EEG channels and one trigger channel indicating events from the ActiveTwoAgent (512 Hz EEG) at 5 Hz (2560 reads/packet). The P300 agent outputted the P300 index at 2 Hz.

The P300 algorithm was fairly complex. It took input from one source, but the input had two components: an EEG data stream and an event indication. There were also two components to the processing: a preprocess component and a detection component. The algorithms were written in MATLAB, so the P300 agent converted the data and passed it into MATLAB, as indicated in Figure C- 2.

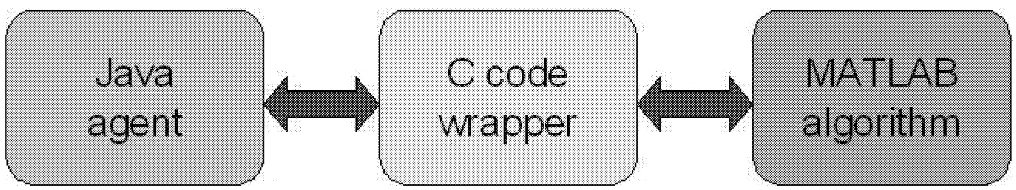


Figure C- 2. P300 language translation.

When the data arrived (SetInputData), three things took place. The first was a check for visual calibration, the second was preprocessing, and the third was detection.

The V-Matrix visual calibration was a task that was done prior to using the gauge and provided a method for removing the noise generated by blinking and eye movement from the EEG data. It was indicated by specific values in the EEG trigger channel. During the visual calibration portion of an experiment, there was a trigger value to indicate the beginning of the task, a trigger for each element of the task (blinking, moving left, moving right, moving up, and moving down), and a trigger to signal the end of the task. During the task, the EEG data was buffered. At the end of the task, the data was passed through the *eyecalibrate.m* algorithm to obtain a V-matrix. The buffered data was saved to a file called *ViscalData.txt* for later analysis, V-matrix was stored in a file called *Vmatrix*, and the preprocess algorithm was reinitialized. If the agent was shut down and restarted, it looked for the *Vmatrix* file and used it for initialization, so there was no need to perform the visual calibration task again.

The preprocessing was accomplished using the *preprocess.m* file. A *preprocessinit.m* file was used to initialize a data structure that built up a history of the EEG data. This data structure was passed into the preprocessing algorithm each time and was stored externally

in a file called *Ppreprocess*. If Vmatrix data was available, it was passed into *preprocessinit.m* for more accurate preprocessing.

The detection was done using the *detect.m* file. There was a *detectinit.m* file that initialized a data structure used to store relevant historical data and improved the accuracy of detection. This structure was stored externally in the *Pdetect* file. Detection was accomplished by taking a 150-millisecond EEG window starting 250 milliseconds after an event. The algorithm then searched this window for a P300 waveform based on positive and negative examples collected during the P300 calibration routine. The gauge calibrated on several examples before it became accurate. Once calibrated, the *Pdetect* file was not updated any further and was used statically to determine the gauge value. When the calibration period ended, a *p300.cal* file was created to indicated completion of the calibration. This file did not contain any data, but was merely used to indicate that the *Pdetect* file was calibrated and did not need to be updated.

During operation, only positive events were sent to the P300 gauge for evaluation. The cost function was adjusted in the *Agent.properties* file to bias the gauge (adjust the false positive/false negative ratio). The default was a 10:1 ratio, biasing toward more false negatives.

C.1.3.5 XLI Gauge

The HBXLoad Index was a measure of executive load or comprehension; positive values indicated increasing load and negative values indicated decreasing load. It operated by measuring power in the EEG at frontal (FCZ) and central midline (CPZ) sites. The algorithm used a weighted ratio of delta + theta/alpha bands calculated during a moving two-second window. The current reading was compared to the previous 20-second running average to determine if the executive load was increasing, decreasing, or staying the same.

The XLI gauge was built as an externally callable .dll under C++ Compiler and was wrapped as a Java Agent for use within the IHMC architecture. For this effort, the XLI was customized by the addition of an internal-calibration module that measures the alpha, beta, delta, and theta frequency bands specific to each participant during a calibration period, and these calculated bands were then used to adjust the XLI configuration files at runtime. The XLI used at IHMC outputs a new value every 2 seconds that was normalized between –1 and 1. The normalization routine used the maximum and minimum raw workload calculations taken from the 20-second running average period to establish the values for the –1 and +1 thresholds. When a new sample was measured, the raw workload value from the XLI algorithm was compared to the maximum and minimum running average and scaled between –1 and 1. At IHMC, the normalized XLI outputs were classified into three ranges—low load, medium load, and high load—by dividing the XLI range in equal thirds. In this manner, the XLI was used to track the differential allocation of executive attentional resources at a rate of 0.5 Hz during a complex divided attention task.

The HBXLoad agent was initialized with the number of input channels (two) and the EEG delta, alpha, beta, and theta values. The HBXLoad agent took as input CPz and FPz

data from the ActiveTwo at 5 Hz (2560 reads per packet). The HBXLoad agent outputted the old XLI index (used for comprehension in the IHMC CVE) and the new XLI index (used as a workload gauge in the CMU CVE) at 1 Hz.

HBXLoad was a MATLAB-based agent. During the visual calibration routine, three executive load tasks were performed (counting backward and reciting the alphabet mentally with eyes open and eyes closed). During this time, the device was calibrated based on the delta, alpha, beta, and theta bands for the individual participant. The XBXLoad agent loaded these bands when restarted prior to each scenario.

Pre-CVE data from both IHMC and CMU were used to enhance the output of the XLI with respect to the attentional bottleneck component being investigated at each site. For instance, an internal calibration mechanism was added to help tune the gauge output based on each person's unique alpha, beta, delta, and theta frequency bands. A trending capability was also added based on preliminary evaluations of the XLI's output with respect to discreet task events. This evaluation led to the development of the rule set used to classify the XLI's ability to differentiate between comprehending or not comprehending an auditory message during an intense continuous-performance task battery.

C.1.4 Practical Constraints and Limitations

Two main practical constraints and limitations were encountered, namely, computational resources and electromagnetic interference (EMI). The former case was addressed by distributing the architecture across multiple PC workstations. The latter issue, EMI, was addressed by selecting a preamplified EEG system (BioSemi ActiveTwo) and careful routing of physiologic device leads. Despite these precautions, significant EMI was still seen in some data files (intermittent and irregular).

C.2 Configuration for CLIP at CMU CVE

C.2.1 Functional Components of the CLIP

The participant was asked to play the part of an upright, mobile military lookout on a virtual rooftop in a simplified urban environment. He or she wore a lightweight, motion-tracked head-mounted display and was given a motion-tracked M16 rifle prop. The gun prop was visible in the virtual environment and produced a red laser dot on objects, indicating precisely where the gun was being aimed. In the environment, four buildings, each in one of the cardinal directions (north, south, east, or west), surrounded the participant. Each building had four columns of evenly spaced windows. The windows of the top four floors on each of the buildings were open, producing a four-by-four array of windows past which friendly or enemy Soldiers could walk. Computer speakers in the room allowed for simulated radio broadcasts to be heard by the participant.

Each participant was outfitted with a BioSemi ActiveTwo EEG cap with 34 scalp electrodes and three ECG electrodes from the UFI EZ-IBI system (two active, one ground). The ActiveTwo and EZ-IBI devices connected to PC workstations via a USB port and serial port, respectively. The physiologic data was captured by the PC

workstations and transferred between agents as needed. The test operator was positioned to monitor the agent and system displays of the PC workstations and could launch and adjust all agents as necessary. Data logging was performed by all agents (sensors and CWA agents and applications) locally in binary form, and the resulting files were collected and posted *post hoc*.

C.2.2 Workstation Configuration

The CMU system was hosted on three desktop computers:

- Panda Desktop ran only Panda-3D simulation.
- EEE-desktop host interfaced with ActiveTwo system with long optical cable coming off the mobile participant and into the USB 2 port of this computer; this computer also ran the architecture agents associated with EEG analysis—ActiveTwo, XLI, and Engagement Index.
- EZ-IBI-desktop host interfaced with EZ-IBI via a long serial cable coming off the mobile participant and into the serial port; this computer also ran the Arousal Meter agents and experimenter console agents such as the mitigation agent.

C.2.3 Sensor/Gauge System Setup

The CMU CVE setup was similar to that described above for the IHMC CVE. Key differences included use of only three gauges (Arousal, Engagement, XLI) in the CMU CVE. Each gauge was connected to specific hardware. Namely, the ActiveTwo (BioSemi, Netherlands) EEG device connected to the HBXload gauge (CPz and FPz), Engagement (P3, P4, Cz, Pz). At CMU, the EZ IBI wearable system was used to provide IBI data, instead of the Cardiax device used at IHMC. Data from the EZ-IBI served as input to the Arousal Meter agent. A primary task baseline period was used to provide the average and standard deviation that is used in generating real-time z-scores for the Engagement Index.

C.2.4 Cognitive State Gauges

Three gauges were used in the CMU CVE: Arousal Meter, Engagement, and XLI.

C.2.4.1 Arousal Meter

The Arousal Meter was functionally equivalent to that of the IHMC CVE, except that at the CMU CVE, the standard desktop Arousal Meter (version 2.2) was used with the recommended EZ-IBI hardware for IBI generation.

C.2.4.2 Engagement Index

The Engagement Index was functionally equivalent to the one in the IHMC CVE, except the version used at CMU did not include custom frequency subbands.

C.2.4.3 XLI

The XLI gauge used at CMU was identical to that used in the IHMC CVE.

C.2.4.4 Practical Constraints and Limitations

During early pilot tasks, it was discovered that some participants could only be immersed for up to 25 minutes before they started experiencing discomfort, such as headaches and mild nausea, associated with prolonged immersion. In addition, the head-mounted eye tracker was not used because of head-mounted display constraints and the availability of participant head space.

Appendix D
Phase 3 CLIP Configuration

The Spring 2005 CVE used a body-worn system that was responsible for all of the sensing, signal processing, reasoning, user interaction management, and data logging.

D.1 Sensor and Mobile Ensemble Deployment

Efforts focused on deployment of the Honeywell team sensor system into a mobile, experiment test environment. The primary challenge was fielding an integrated sensing, computational and interactive system within a mobile hardware ensemble. The prototype ensemble was organized around the US Army MOLLE backpack that provided the framework on which to integrate multiple sensors, interface devices, network adapters, and the data collection computer.

Transitioning from a laboratory environment with computer simulations to a field exercise required network communications to support experiment task support such as scripting and stimuli presentation. For the field experiments, a remote computer ran the scripts that played pre-recorded radio broadcasts to simulate communication traffic to a dismounted infantry leader. Initially, all sensed data were transmitted wirelessly to a remote desktop computer that reasoned on cognitive state to trigger mitigations and also logged data for post hoc analysis.

Network connectivity and reliability across the experiment test field was a considerable challenge and motivated the migration of all data logging and reasoning to the backpack laptop; moreover, even on-body network communications from the UFI/Clemson Wearable Arousal Meter (WAM) and the Anthrotronix Tactabelt proved to be unreliable enough to require a reversion back to wired connections to improve performance. After streamlining the EEG signal conditioning algorithms, migrating all hardware interfaces to the backpack laptop, and integrating and testing the Point Research (now Honeywell International, Inc.) GyroDRM module, an early system integration test was performed for a technical progress meeting with the Army and DARPA in November 2004. Subsequently, all software components for signal processing, adaptive system (mitigation) reasoning, and data logging were migrated to the backpack computer. Following a successful system integration test, a full evaluation was conducted in December 2004 that included both semi-mobile and fully mobile multi-tasking scenarios with operational relevance to the Army.

Early in 2005, the InertiaCube head tracker and a Pocket PC device were integrated for use in the Spring 2005 CVE. Later that spring ABM's 6-channel EEG system, which communicated wirelessly via Bluetooth from the sensor headset to the backpack computer, was integrated and tested. After integrating the GyroDRM data stream with the Tactabelt for use as a navigation-aiding mitigation, a final system integration test was conducted for an evaluation conducted in the spring. This integration included hardware and software for ABM EEG, Clemson/UFI Arousal Meter, Intersense Headtracker, Point Research Dead Reckoning Module, head-mounted Web-cam, Pocket PC-based distraction task, Pocket-PC based Communications Scheduler mitigation, and GyroDRM

& Tactabelt-based navigation-aiding mitigation—all run on a single laptop stowed in a backpack communicating over an ad hoc and Bluetooth networks (see Figure D-1).

Figure D-1. The Honeywell mobile ensemble, used in the Spring CVE.

D.2 Description of CLIP

The mobile ensemble was integrated in a modified US Army MOLLE system (Modular Lightweight Load Bearing Equipment Fully compatible with U.S. Military Style), as pictured in Figure D- 2. The integrated hardware/software solution supported sensing and user interaction. A modular agent architecture supported data integration, signal processing, reasoning, and data logging.

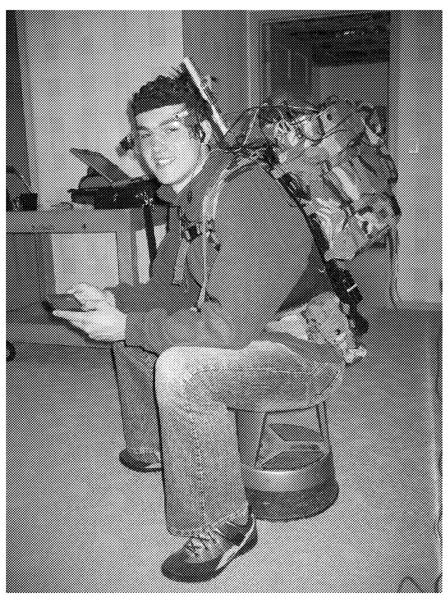


Figure D- 2. The mobile ensemble integrated in Army MOLLE system.

The integrated hardware solution comprised of the following systems:

- ABM's six electrode electroencephalogram (EEG) system measured cognitive brain activity;
- UFI/Clemson's Cognitive Wearable Arousal Meter measured the heart's interbeat interval as an index of alertness;

- Point Research's Dead Reckoning Module provided bearing, activity levels, and location information based on integrated accelerometers and global position system;
- InterSense's head tracker provided head orientation and movement information as an index of visual attention in the field;
- Anthrotronix's tactabelt for vibrotactile cueing; a commercial Web-cam that delivered line-of-sight video capture; and
- Hewlett Packard's iPAQ PocketPC displayed messages and data logging during demonstrations and exercises.

Hardware input and data logging was managed by the agent-based information architecture developed in concert with the Institute for Human and Machine Cognition (IHMC). The current architecture enabled the components of cognitive state assessment such as hardware sensors (e.g., EEG, ECG) and software algorithms to be integrated and tested.

This architecture (see Figure D- 3) supported end-to-end reasoning that assessed cognitive state and detected context in order to select an appropriate mitigation response. In addition, the architecture provided experimental and data management support within a common logging format.

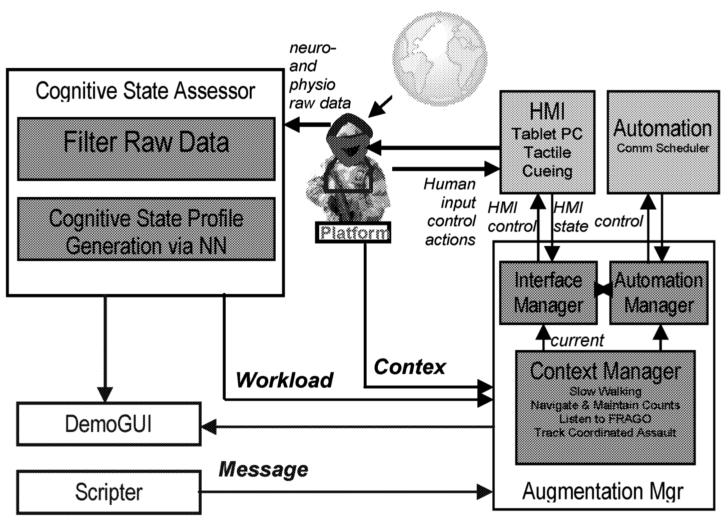


Figure D- 3. The CLIP architecture.

D.3 System Components

The system consisted of the following component classes:

- Physiological Sensors
- Context Sensors
- Hardware Interface Agents: managed data interface between hardware and software architecture
- User interaction devices: serve as input and output devices for participants

- Load carriages system: modified backpack that holds all devices
- Computing platform: laptop and peripheral devices that support port expansion and network connectivity
- Signal processing agents: software agents that condition EEG signal
- Experiment Management and Data Logging Processes: software agents that manage experiment execution and data logging

Physiological Sensors

- ABM EEG: 6 channels of Raw EEG, decontaminated power spectral densities in 1-Hz bins, ABM workload gauge, ABM vigilance gauge
- UFI/Clemson WAM (ECG): Arousal meter and interbeat interval (IBI)

Context Sensors

- Point Research GyroDRM: mounted on backpack frame to derive the following: Activity level, location, bearing, reconstructed participant path
- InterSense InertiaCube: mounted on safety glasses worn by participant to record head movement (yaw, pitch, roll)
- WebCam: mounted on safety glasses worn by participant to record point of regard video

Hardware Interface Agents: manages data from sources

- ABM_EEG_Raw
- WAM
- GyroDRM
- InertiaCube
- Video_Capture_Agent

User interaction Devices

- Radios: medium for situational awareness; participants respond via radio; supports Mission Monitoring and Counts secondary tasks
- HP iPAQ PDA: Math interruption task completed on PDA; supported mitigated performance of Counts task—counts communications deferred, in text, to be reviewed on PDA
- AnthorTronix Tactabelt: vibrotactile input belt used in navigation support during which the direction to the next waypoint buzzes

Load carriage System

- Modified US Army MOLLE backpack
- Cooling fans: to ensure that laptop does not overheat in backpack

Computing Platform

- Dell D600 Laptop
- USB Battery-powered Hub
- Bluetooth Network: to support wireless communications with ABM headset

Signal Processing Agents

- ABM_PSD: samples power spectral density in 1-Hz bin from ABM system; display running average of spectral power used in identifying noise prior to experiment runs
- ABM_PSD_Combiner: combines 1-Hz bin into clinical bands
- ABM CWPC: samples ABM proprietary gauges: workload and vigilance
- Cognitive_State_Classifier: support training and real-time detection of cognitive state based on a trained classifier using spectral power as input

Experiment Management and Data Logging Processes

- Agent Launcher: user interface to start all agents
- Qscripter & FieldScripter: enables playing script of digitally recorded audio message to support Mission Monitoring, Counts, and Math interruption
- Experimental Console: provides feedback about state of agents (logging, active, error-condition) and enables starting and stopping data logging